Forecasting Financial Returns Under Non-Elliptical Distributions with Applications to Portfolio Allocation and Risk Management

Dissertation
submitted to the Faculty of Economics,
Business Administration and Information Technology
of the University of Zurich

to obtain the degree of
Doctor of Philosophy
in Banking and Finance

presented by

Paweł Polak
from Poland

approved in December, 2013 at the request of
Prof. Dr. Marc S. Paolella
Prof. Loriano Mancini, Ph. D.
Prof. Dr. Erich Walter Farkas

1. Reviewer: Prof. Dr. Marc S. Paolella

2. Reviewer: Prof. Loriano Mancini

3. Reviewer: Prof. Dr. Erich Walter Farkas

Day of the defense: December 4th, 2013

Signature from head of PhD committee:

To my parents and Anna.

Acknowledgements

I would like to thank all my colleagues, friends, and family who contributed to this thesis! Special thanks goes to my doctoral advisor professor Paolella, with whom I worked together very closely, had numerous discussion, and excellent brainstorming sessions which lead to this dissertation and to a pipeline full of projects on which we continue working. I would like to thank Jochen Krause for discussions over numerical analysis, optimization methods, and Matlab programming. Last but not least, I want to thank my parents and Anna for their constant support.

Declaration

I herewith declare that I have produced this paper without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This paper has not previously been presented in identical or similar form to any other Swiss or foreign examination board.

The thesis work was conducted under the supervision of Prof. Dr. Marc S. Paolella at the University of Zurich.

Zurich, December 4, 2013

Contents

List of Figures

List of Tables

1

Introduction

After numerous financial crashes of different scales which have been observed in recent years, the issue of poor risk management practices has been raised by many researchers and practitioners. In particular, the assumption of normality of asset returns has been named as one of the sources of a non-negligible model risk which has led to substantial losses of financial institutions in which long term risk management is crucial, e.g., banks, insurance companies and pension funds. As Greenspan (1997, p. 54) states,

> ... the biggest problems we now have with the whole evaluation of risk is the fat-tail problem, which is really creating very large conceptual difficulties. Because as we all know, the assumption of normality enables us to drop off a huge amount of complexity of our equations very much to the right of the equal sign. Because once you start putting in non-normality assumptions, which unfortunately is what characterizes the real world, then the issues become extremely difficult.

The dissertation consists, of a series of three papers, with new multivariate, non-elliptical models for asset returns, such as stocks, bonds, and currencies, which (i) are able to predict the entire multivariate density function; (ii) can do so using hundreds or even thousands of assets, on conventional computing machines with standard software; and (iii) are such that the corresponding portfolio distribution, which is necessary to evaluate advanced risk measures such as value at risk (VaR) and expected shortfall (ES), is analytically tractable. These papers provide also the necessary tools for portfolio optimization, risk management, and option pricing, for the proposed models, and empirically confirm the in-sample fit and the out-of-sample superiority of the new models to numerous competing ones.

Chapter 2 motivates the use of the multivariate generalized hyperbolic (MGHyp) distribution which forms the building block of all the proposed models.

Chapter 3, consists of the first paper, coauthored with Marc S. Paolella, in which we consider modeling a set of asset returns via a conditional multivariate distribution with dynamics governed by a process which has features of both GARCH and stochastic volatility (SV). These two essentially disparate paradigms for capturing volatility clustering in asset returns have their individual advantages, and also limitations. Our model is a judicious hybrid which utilizes only advantages of the two dynamics. To the best of our knowledge, this is the first volatility model which combines GARCH and SV paradigms—and is applicable to large-dimensional multivariate settings, owing to the proposed EM-algorithm for maximum likelihood estimation (MLE) and the possibility for parallel computing. Estimation is numerically reliable and fast, and can be used even with a large number of assets, crucially, yielding consistent and asymptotically normal estimates of the parameters. An extensive empirical exercise involving the 30 assets comprising the DJIA reveals that, in terms of in-sample fit and out-of-sample density forecasts based on the predictive log likelihood, the new model demonstrably outperform the classical CCC model of Bollerslev (1990), the VC model of Tse and Tsui (2002), the DCC model of Engle (2002), and the cDCC model of Aielli (2011). As such, the proposed model can be used for deriving an accurate prediction of the conditional mean vector and covariance matrix of a set of asset returns, and so has a direct use in market risk management. In addition, because of the tractability of the sums of margins, portfolio allocation based on minimization of downside risk measures such as the value at risk and expected shortfall can be conducted, as well as based on the mean and variance, if desired. Finally, by combining the equivalent martingale measure approach with Monte Carlo simulation, option pricing can also be conducted.

Chapter 4 consists of a second, single authored paper, which is motivated by the stylized fact that the dependency between assets is not constant over time. The paper proposes a new non-elliptical model which utilizes the aforementioned correlation dynamics. The main advantages of these models are their applicability to a large number of assets, ease of estimation, and improved model fit compared to a CCC model of Bollerslev (1990). However, for multivariate density forecasting, risk forecasting, and portfolio allocation, a severe disadvantage of these models is the underlying assumption of Gaussianity for the innovation sequence. We propose a model which utilizes a (possibly special case of a) MGHyp distribution, which accounts for asymmetry and excess kurtosis in the returns, and, like the Gaussian models, has a tractable sums of marginals (as required for portfolios). Another novelty of the model is the estimation

of all its parameters by a joint maximum likelihood method, using an EM algorithm, so that it is still feasible for hundreds of assets. Empirical results demonstrate that the new models provide improved density forecasts, better risk predictions, and higher performance of expected-shortfall-optimized portfolios than aforementioned Gaussian models.

Since the seminal work of Hamilton (1989), Markov switching models constitute a prominent role in financial econometrics. Chapter 5, consists of the third, single authored, paper. It is dedicated to a regime-switching model for a multivariate set of asset returns which extends the CCC model of Bollerslev (1990) and the regime switching model of Pelletier (2006), to a (possibly special case of a) MGHyp. Crucially for portfolio applications and risk predictions, the new model maintains tractability of the sums of marginals. A new, two-stage, EM algorithm is developed for estimation and the asymptotic properties of the corresponding parameters estimator are investigated. In order to enhance the forecast quality, the model is coupled with shrinkage via a quasi-Bayesian prior, and it is demonstrated to outperform all special cases in terms of in sample fit and out-of-sample forecasts.

2

Multivariate Generalized Hyperbolic Distribution

Part of the appeal of using the normality assumption is model simplicity and ease of parameter estimation. Moreover, it is known that, under certain assumptions on the true data generating process, the method of quasi maximum likelihood (use of normality even though it is incorrect) still leads to a consistent estimate of the covariance matrix or, more generally, the parameters which govern its evolution through time (see, e.g., Francq and Zakoïan, 2004, 2010). This approach, however, can suffer in small samples because of the divergence of the Gaussian and true innovations distribution. In particular, a large deviation can cause an increase in the variance of the estimates, and result in efficiency loss. This is particularly acute in empirical finance applications, for which asset returns data are well-known to be generally heavy-tailed and mildly asymmetric, and even more so in the common case with a large number of assets compared to observations in time; see, e.g., Bickel and Levina (2008); Fan et al. (2008); and the references therein.

Furthermore, if interest centers not just on asymptotically consistent point estimates of some of the model parameters, but rather on the predictive distribution of the returns based on a potentially small finite sample, then the incorrect use of normality may have far-reaching consequences. In particular, modern quantities for portfolio risk, such as value-at-risk and expected shortfall, do not just depend on the mean vector and covariance matrix of returns, but on all the distributional parameters. Such risk measures are features of the predictive distribution, which is the most general object from which all measurable quantities of interest can be derived. Indeed, in both the univariate case (see, e.g., Kuester et al., 2006b; Broda

et al., 2012; and the references therein) and multivariate case (see, e.g., Paolella, 2013), risk and density forecasts are greatly enhanced when moving from the normal to flexible, heavier-tailed, asymmetric distributions. Observe also that portfolio allocation which respects the need for downside-risk measures and the non-normality of asset returns requires the use of non-Gaussian distributions; see, e.g., Doganoglu et al. (2007); Giacometti et al. (2007); and Broda and Paolella (2009).

It is important to emphasize that the true underlying class of distributions need not be known, but only that the assumed distributional class is flexible enough to capture the salient features of the data (in finance, these being heavy tails and asymmetry, but otherwise a uni-modal, bell-like distribution which nests, or features as a limiting case, the normal distribution). The application of a parametric model which most likely differs from the true underlying one, but is flexible enough to adequately fit the data, is generally referred to as *partially adaptive estimation*; see, e.g., McDonald and Newey (1988); Phillips (1994); and Hansen et al. (2006). It is useful because, besides accounting for the often substantial departures from normality, it avoids some of the disadvantages of nonparametric inference (see, for example, McDonald, 1991, 1997), even if, as already mentioned, the underlying model is not precisely from the specified distributional class. The MGHyp distribution is essentially perfect for this task, given its great flexibility, but also its other desirable features, as discussed next.

The MGHyp was introduced in Barndorff-Nielsen (1978) and popularized as a candidate model for financial returns in Eberlein and Keller (1995), Eberlein et al. (1998), and McNeil et al. (2005). It is a very general and extremely flexible (albeit uni-modal) class of distributions which, as special or limiting cases, includes asymmetric versions of the (multivariate) Student's t, Laplace, normal inverse Gaussian, variance-gamma, Gaussian, and others; see Paolella (2007, Ch. 9) for a detailed presentation in the univariate case. An interesting feature of this distribution which is particularly relevant for financial returns data is that it has what are called semi-heavy tails; these being a compromise between genuine power (fat) tails and exponential (thin) tails. Moreover, the special and limiting cases mentioned above indicate that it also supports (in the limit) both fat and thin tails. Theoretical support for using an MGHyp model comes from the fact that it is consistent with continuous-time models where logarithmic asset prices follow multivariate Lévy processes; see, e.g., Eberlein and Keller (1995). Another theoretical property which is of great importance for portfolio optimization is that, if the vector of asset returns is MGHyp distributed, then the distribution of the portfolio return is also GHyp with location and skewness parameters being weighted (by portfolio weights)

sums of the corresponding vector parameters and the scale parameter being a quadratic form of the corresponding dispersion matrix. In particular, using the notation introduced below, if $\mathbf{Y} \sim \text{MGHyp}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{H}, \lambda, \chi, \psi)$, then $\mathbf{w}'\mathbf{Y} \sim \text{GHyp}(\mathbf{w}'\boldsymbol{\mu}, \mathbf{w}'\boldsymbol{\gamma}, \mathbf{w}'\mathbf{H}\mathbf{w}, \lambda, \chi, \psi)$ for any vector $\mathbf{w} \in \mathbb{R}^K \setminus \mathbf{0}$, where K is the dimension of $\mathbf{Y}$, and the parameter spaces for $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$, $\mathbf{H}$, λ, χ and ψ are given below.

From an empirical perspective, it has been shown (for both univariate and multivariate data) that the MGHyp provides an excellent fit to the unconditional distribution of financial returns; see, e.g., McNeil et al. (2005) and the reference therein. (Though note that an unconditional model cannot capture the typical stylized facts of time-varying variances and correlations; this is dealt with herein.) However, the extension from normality to MGHyp entails estimation of yet more model parameters. Far worse, however, is that, in high dimensions (irrespective of the distribution), estimation via direct optimization of the likelihood becomes problematic because of the proliferation of parameters in the dispersion matrix and, for our conditional models below, the GARCH parameters governing the time-varying evolution of each of the univariate processes. To address this, a multi-stage Expectation Maximization (EM) algorithm is developed for maximum likelihood estimation. This is possible because the MGHyp is expressible as a continuous normal mixture; see below for details. Crucially, the new estimation method is applicable for a large number of assets.[1]

Finally, by the mixture structure of the MGHyp, the univariate mixing component ($G_t \sim \text{GIG}(\lambda, \chi, \psi)$; see below) can be treated as a common market factor which is amenable to prediction, and is key to allowing the model to embody an SV extension.

[1] Blæsid and Sørensen (1992) developed a program for estimating MGHyp distributions, for which Prause (1999) reports that it can estimate the general case up to three dimensions in a reasonable time. An EM algorithm was proposed by Protassov (2004). All this work was for iid data, and does not include any type of multivariate conditional volatility modeling. Mencía and Sentana (2009) propose an ML estimator for MGHyp in a conditional setup. Their estimation is based on the analytical expressions for the log-likelihood score, which improves its performance, but it does not solve the curse of dimensionality problem.

3

COMFORT

COMFORT: A Common Market Factor Non-Gaussian Returns Model

Marc S. Paolella[a,b] Paweł Polak[a,b]

[a] *Department of Banking and Finance, University of Zurich, Switzerland*
[b] *Swiss Finance Institute*

Abstract

A new multivariate time series model with various attractive properties is motivated and studied. By extending the CCC model in several ways, it allows for all the primary stylized facts of financial asset returns, including volatility clustering, non-normality (excess kurtosis and asymmetry), and also dynamics in the dependency between assets over time. A fast EM-algorithm is developed for estimation. The predictive conditional distribution is a (possibly special case of a) multivariate generalized hyperbolic, so that sums of marginals (as required for portfolios) are tractable. Each element of the vector return at time t is endowed with a common univariate shock, interpretable as a common market factor, and this stochastic process has a predictable component. This leads to the new model being a hybrid of GARCH and stochastic volatility, but without the estimation problems associated with the latter, and being applicable in the multivariate setting for potentially large numbers of assets. Its applicability to option pricing is developed. In-sample fit and out-of-sample conditional density forecasting exercises using daily returns on the 30 DJIA stocks confirm the superiority of the model to numerous competing ones.

Keywords: CCC; Common Jumps; Density Forecasting; EM-Algorithm; Fat Tails; GARCH; Multivariate Asymmetric Laplace Distribution; Multivariate Asymmetric t Distribution; Multivariate Generalized Hyperbolic Distribution; Multivariate Normal Inverse Gaussian Distribution; Multivariate Option Pricing; Stochastic Volatility.

JEL Classification: C51; C53; C58; G11; G17.

3.1 Introduction

We consider modeling a set of asset returns via a conditional multivariate distribution with dynamics governed by a process which has features of both GARCH and stochastic volatility (SV). These two essentially disparate paradigms for capturing volatility clustering in asset returns have their individual advantages, and also limitations. In univariate models, both can capture changes in volatility, the leverage effect (at least for more recent SV models) and the leptokurtosis and possible asymmetry of the innovations distribution. The main disadvantage of SV models is the lack of an explicit form of the likelihood function and the necessity to

use moment-based methods or simulation for estimation. The former are often simple, but inefficient, while the latter attempt to achieve a close approximation of the likelihood function through computationally intensive methods, and become problematic for more than a small number of assets (see, e.g., Asai et al. 2006; Asai and McAleer 2009; and Bos 2012; and the reference therein). Alternatively, in the univariate case, a GARCH model is trivial to estimate via (conditional) maximum likelihood, but is unable to model the dynamics of an unpredictable volatility component. Both SV and many kinds of GARCH models share the same estimation infeasibility in the large-scale multivariate case, though models such as CCC, DCC, and some of their extensions, are feasible; these are considered in the empirical section below.

The recent literature has emphasized the importance of including a stochastic jump component in the volatility dynamics—a feature which can be easily incorporated into the SV structure (see, among others, Chernov et al. 2003; Eraker et al. 2003; Eraker 2004; and Todorov and Tauchen 2011) but is absent in GARCH models. (One exception is the model proposed by Chan and Maheu 2002, and Maheu and McCurdy 2004, see also Section 4.2 below.) Another approach is to use high-frequency returns in order to construct realized variation measures which separate the volatility jump component from the smooth continuous movement of the underlying volatility (see, e.g., Andersen et al. 2007).

In the multivariate case, jumps in individual assets can be observed to arrive simultaneously, forming what are called co-jumps. Following the arbitrage pricing theory, Bollerslev et al. (2008) distinguish the co-jumps in the idiosyncratic component, which are diversifiable, and the co-jumps in the common component, which are non-diversifiable, i.e., those which carry a risk premium. According to their statistical test, there are co-jumps which can be highly significant, but when considered on individual stocks, remain undetected. The importance of co-jumps structures for the consumption-portfolio selection problem is discussed by Aït-Sahalia et al. (2009). More recently, Gilder et al. (2013) provide empirical evidence for co-jumps and analyse the association between jumps in the market portfolio and co-jumps in the underlying stocks. Their results suggest that news events which have market-level influence are able to generate large co-jumps in individual stocks, while Bollerslev et al. (2013) show that extreme joint dependencies observed in daily data can be implied by the diffusive volatility *and* co-jumps observed in the high-frequency data.

In line with this high-frequency literature, we propose a new model which also splits the dynamics of volatility; however, it is a parametric approach which does not require high-frequency data and is applicable in the multivariate setting. We propose a solution which, in a conditional setup, utilizes a flexible, fat-tailed distribution, and combines univariate GARCH-type dynamics (most of the popular variations are possible) with a relatively simple, yet flexible, SV dynamic structure, based on the seminal work of Taylor (1982). By introducing a latent component similar to that used in the SV literature, the resulting hybrid GARCH-SV model is able to capture stochastic (co-)jumps in the volatility series and across assets. To the best of our knowledge, this is the first volatility model which combines GARCH and SV paradigms—and is applicable to large-dimensional multivariate settings, owing to the proposed EM-algorithm for maximum likelihood estimation (MLE) and the possibility for parallel computing.

The model contains a univariate, latent component which is common to all K assets. We term this a common market factor. It dictates the nonlinear dependency between margins, so that, conditional upon it (as realized via the EM-algorithm), what remains to be estimated are K univariate normal GARCH models. The (conditional) MLE of each of the latter is numerically fast and reliable to obtain, and the K estimations can be conducted in parallel. We show that the latent common market factor process has a predictable component, and that this leads to (statistically significant) improvements in forecasting performance over and above the sizeable improvements obtained by relaxing the normality assumption.

Another important feature of our model is that it imposes a multiplicative structure on the volatility, and an infinitely divisible distribution which, in the iid case, generates an infinite activity jump process, as opposed to jump diffusion models, where finite activity jumps, modeled via a Poisson-distributed term, are added to Gaussian dynamics. This is in line with Aït-Sahalia and Jacod (2011), who propose two non-parametric statistical tests to discriminate between the two cases, and both tests point toward the presence of infinite-activity jumps in the data. Other examples of models which support infinite activity jumps, and are often used in the context of option pricing, are the variance gamma model of Madan and Seneta (1990) and Madan et al. (1998), the hyperbolic model of Eberlein and Keller (1995), and the CGMY model of Carr et al. (2002).

An extensive empirical exercise involving the 30 assets comprising the DJIA reveals that, in terms of in-sample fit and out-of-sample density forecasts based on the predictive log likelihood, the new model demonstrably outperforms the classical CCC and DCC models. As such, the proposed model can be used for deriving an accurate prediction of the conditional mean vector and covariance matrix of a set of asset returns, and so has direct use in market risk management. In addition, because of the tractability of the sums of margins, portfolio allocation based on minimization of downside risk measures such as the value at risk and expected shortfall can be conducted, as well as based on the mean and variance, if desired. The general model, with parameter $\boldsymbol{\gamma}$ nonzero, is non-elliptic, and the data support $\boldsymbol{\gamma} \neq \mathbf{0}$, so that use of coherent risk measures, such as expected shortfall, in portfolio optimization, will have a positive effect on the management of portfolio risk; see Embrechts et al. (2001). Finally, by combining the equivalent martingale measure approach with Monte Carlo simulation, option pricing can also be conducted.

The remainder of the paper is as follows. The model and some of its properties are stated in Section 4.2. Section 4.3 discusses the proposed method of estimation. The MGHyp distribution being essentially too flexible, Section 3.4 motivates and details important special cases. Section 3.5 provides details on option pricing. Section 4.5 illustrates an exercise with real data in which in-sample fit and out-of-sample density forecasting are compared across models. Section 3.7 provides some concluding remarks, and an Appendix gathers various technical results.

3.2 Model

We consider a set of K financial assets, with associated return vector at time t given by $\mathbf{Y}_t = (Y_{t,1}, Y_{t,2}, \ldots, Y_{t,K})'$, $t = 1, 2, \ldots, T$, whose conditional, time-varying distribution is taken to be multivariate generalized hyperbolic, hereafter MGHyp; see, e.g., McNeil et al. (2005). We observe a realization of $\mathbf{Y} = [\mathbf{Y}_1 \mid \mathbf{Y}_2 \mid \cdots \mid \mathbf{Y}_T]$, where the $\mathbf{Y}_t$ are equally spaced in time (ignoring the weekend effect for daily data) return vectors. The information set at time t is currently defined as the history of returns, $\mathbf{\Phi}_t = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_t\}$, though extensions to the model which include the use of exogenous variables could be straightforwardly entertained. Our focus is on the prediction of the conditional distribution of $\mathbf{Y}_{T+h}$ given $\mathbf{\Phi}_T$, for which we restrict our empirical demonstration in this paper to $h = 1$. The dispersion matrix of $\mathbf{Y}_t \mid \mathbf{\Phi}_{t-1}$ is decomposed as the product of scale terms and a conditional dependency matrix (a correlation matrix when the MGHyp distribution approaches the multivariate normal). For each of the univariate scales, a GARCH-type structure is imposed, while the dependency matrix is specified as being constant over time. We will see in Remark 4.2.(ii) below that, except for some special cases, the correlations are actually time-varying. Further methods of inducing time-variation in the dependency matrix are discussed in the conclusions.

Using the mixture representation of the MGHyp (see Section 4.3 below; Eberlein and Keller, 1995; and Eberlein et al., 1998), we can express the return vector as

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\gamma} G_t + \boldsymbol{\varepsilon}_t, \quad \text{with} \tag{3.1a}$$

$$\boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2} \sqrt{G_t} \mathbf{Z}_t, \tag{3.1b}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$ are column vectors in $\mathbb{R}^K$; $\mathbf{H}_t$ is a positive definite, symmetric, conditional dispersion matrix of order K; $\mathbf{Z}_t \overset{\text{iid}}{\sim} \mathrm{N}(\mathbf{0}, \mathbf{I}_K)$ is a sequence of independent and identically distributed (iid) normal random vectors and $(G_t \mid \mathbf{\Phi}_{t-1}) \sim \mathrm{GIG}(\lambda_t, \chi_t, \psi_t)$ are mixing random variables, $t = 1, 2, \ldots, T$, independent of $\mathbf{Z}_t$, with typical GIG (generalized inverse Gaussian) density given by

$$f_G(x; \lambda, \chi, \psi) = \frac{\chi^{-\lambda} \left(\sqrt{\chi\psi}\right)^{\lambda}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} x^{\lambda-1} \exp\left(-\frac{1}{2}\left(\chi x^{-1} + \psi x\right)\right), \quad x > 0; \tag{3.2}$$

K_λ is the modified Bessel function of the third kind (and not to be confused with K, the number of assets), given by

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty t^{\lambda-1} \exp\left(-\frac{x}{2}\left(t + t^{-1}\right)\right) \mathrm{d}t, \quad x > 0; \tag{3.3}$$

and $\chi_t > 0$, $\psi_t \geq 0$ if $\lambda_t < 0$; $\chi_t > 0$, $\psi_t > 0$ if $\lambda_t = 0$; and $\chi_t \geq 0$, $\psi_t > 0$ if $\lambda_t > 0$.

We consider two specifications for the GIG parameters: (i) $G_t \mid \mathbf{\Phi}_{t-1}$ are iid with time-invariant parameters, i.e., $\lambda_t = \lambda$, $\chi_t = \chi$ and $\psi_t = \psi$; (ii) $G_t \mid \mathbf{\Phi}_{t-1}$ has time dependent, random, parameters with the dynamics described by a system of conditional moment equations

$$\mathbb{E}\left[G_t^r \mid \mathbf{\Phi}_{t-1}\right] = c_r + \rho_r \mathbb{E}\left[G_{t-1}^r \mid \mathbf{\Phi}_{t-2}\right] + \zeta_{r,t}\,, \tag{3.4}$$

for a set of positive integer values of r; $\zeta_{r,t} = \mathbb{E}[G_t^r \mid \mathbf{\Phi}_t] - \mathbb{E}[G_t^r \mid \mathbf{\Phi}_{t-1}]$; and c_r and ρ_r are parameters to be estimated. The error term $\zeta_{r,t}$ represents the unpredictable component affecting the rth moment of the mixing variable G_t. It contains all new information in forming expectations about G_t^r when moving from time $t-1$ to t. It can be used as a driver of the dynamics of $\mathbb{E}[G_t^r \mid \mathbf{\Phi}_{t-1}]$ in (4.4) because it is a martingale difference sequence (MDS) with respect to $\mathbf{\Phi}_{t-1}$, implying that $\mathbb{E}[\zeta_{r,t}] = 0$ and $\mathrm{Cov}(\zeta_{r,t}, \zeta_{r,t-s}) = 0$, $s = 1, 2, \ldots$. From (4.4), the dynamics of the parameters λ_t, χ_t, and ψ_t can be inferred by the expression for the moments of the GIG random variable given below in (4.6). However, for each of the three special cases of the MGHyp distribution we entertain, it turns out that the dynamics of only one of the three parameters associated with G_t needs to be modeled; see Remark 4.2.(v) below, and Section 3.4, for details. For example, when using $r = 1$ in the estimation of the dynamics in (4.4), $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$ has to be positive. The dynamics in (4.4) can be rewritten as

$$\mathbb{E}[G_t^r \mid \mathbf{\Phi}_{t-1}] = \frac{c_r}{2} + \frac{\rho_r}{2}\mathbb{E}\left[G_{t-1}^r \mid \mathbf{\Phi}_{t-2}\right] + \frac{1}{2}\mathbb{E}[G_t^r \mid \mathbf{\Phi}_t], \tag{3.5}$$

so that the sufficient condition for $\mathbb{E}[G_t^r \mid \mathbf{\Phi}_{t-1}]$ to be positive for all $t > 0$ is $c_r > 0$ and $\rho_r \geq 0$.

When the parameters of G_t are allowed to be dynamic as in (4.4), we call model (4.1) a hybrid GARCH-SV extension because it can be linked to the seminal SV model of Taylor (1982); this link being detailed in Appendix 3.8.1. Our model differs from it because ours is a multivariate model with GARCH dynamics in the individual scales and an SV component which describes the dynamics of the common market factor G_t (as detailed below). Moreover, the dynamics in (4.4) are in the same vein as Chan and Maheu (2002) and Maheu and McCurdy (2004), who model dynamics of the jump intensity of a Poisson process in individual stock returns. (One important difference is that our dynamics imply that G_t^r is *not* a deterministic function of the past returns.) In line with these works, we model the dynamics of the linear projections of G_t^r on past returns only, this being another difference between our approach and that of Taylor (1982).

The dynamics in (4.4) remind an AR(1) process because the output variable depends linearly on its own previous values. Hence, similarly to AR(1) process, and due to the MDS property of the $\zeta_{r,t}$ innovations, the conditional forecasts of the future conditional moments are given by

$$\mathbb{E}\left[G_{t+s}^r \mid \mathbf{\Phi}_t\right] = c_r \sum_{i=0}^{s-1} \rho_r^i + \rho_r^s \, \mathbb{E}[G_t^r \mid \mathbf{\Phi}_{t-1}], \quad s \geq 1, \tag{3.6}$$

where $\mathbb{E}[G_t^r \mid \mathbf{\Phi}_{t-1}]$ is measurable with respect to the information up to time $t-1$ and is given by (4.6) below. If $|\rho_r| < 1$, then the process in (4.4) is mean-reverting, and for $s \to \infty$, the forecast approaches the unconditional mean value $c_r/(1-\rho_r)$ of G_t^r.

The conditional dispersion matrix $\mathbf{H}_t$ is decomposed as

$$\mathbf{H}_t \equiv \mathbf{S}_t \mathbf{\Gamma} \mathbf{S}_t, \tag{3.7}$$

where $\mathbf{S}_t$ is a diagonal matrix composed of the strictly positive conditional scale terms $s_{k,t}$, $k = 1, 2, \ldots, K$, and $\mathbf{\Gamma}$ is a time-invariant, symmetric, with ones on the main diagonal, conditional

dependency matrix, such that $\mathbf{H}_t$ is positive definite. The univariate scale terms $s_{k,t}$ are modeled by a GARCH-type process. In particular, the simplest realistic choice is the GARCH(1,1) model

$$s_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{k,t-1}^2 + \beta_k s_{k,t-1}^2, \tag{3.8}$$

where $\varepsilon_{k,t} = y_{k,t} - \mu_k - \gamma_k G_t$ is the kth element of the $\boldsymbol{\varepsilon}_t$ vector in (4.1), and $\omega_k > 0$, $\alpha_k \geq 0$, $\beta_k \geq 0$, for $k = 1, 2, \ldots, K$. More general formulations could be used, notably those which can capture an asymmetry effect; see, e.g., Mittnik and Paolella (2000) and the references therein. For this purpose, we consider the GJR-GARCH$(1, 1)$ model of Glosten et al. (1993), given by

$$s_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{k,t-1}^2 + \eta_k \varepsilon_{k,t-1}^2 \mathbf{1}\{\varepsilon_{k,t-1} > 0\} + \beta_k s_{k,t-1}^2, \tag{3.9}$$

where $\mathbf{1}\{\cdot\}$ is an indicator function, and $\eta_k \geq 0$ captures asymmetry in the scale-term response to the last period innovation. Engle and Ng (1993) have shown that, in the Gaussian case, the GJR-GARCH$(1, 1)$ model is the best performing parametric model for capturing the asymmetry response of the volatility to news.

Remarks:

(i) In order to maintain the news effect in future volatilities, the innovation process used in the GARCH recursions in (4.8) and (4.9) is $\varepsilon_{k,t} = y_{k,t} - \mu_k - \gamma_k G_t$. If instead, we were to use $\varepsilon_{k,t}/\sqrt{G_t}$, then the next period volatility would not be influenced by the current spike in G_t. Hence, there would be no volatility persistence after news; the model would not capture the stylized fact of volatility clustering, and use of the GARCH-type dynamics for the scale term would be inadequate. We have also empirically confirmed that this alternative specification leads to a very low-persistence GARCH recursion (in terms of $\alpha_k + \beta_k$ values) and that the forecasting performance of the model substantially decreases.

(ii) In model (4.1), $\boldsymbol{\mu}$ is the location vector and $\mathbf{H}_t$ is the dispersion matrix of the conditional distribution of $\mathbf{Y}_t$, while the mean and the covariance matrix are given by

$$\mathbb{E}\left[\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1}\right] = \boldsymbol{\mu} + \mathbb{E}\left[G_t \mid \boldsymbol{\Phi}_{t-1}\right] \boldsymbol{\gamma} \tag{3.10}$$

and

$$\mathrm{Cov}\left(\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1}\right) = \mathbb{E}\left[G_t \mid \boldsymbol{\Phi}_{t-1}\right] \mathbf{H}_t + \mathbb{V}\left(G_t \mid \boldsymbol{\Phi}_{t-1}\right) \boldsymbol{\gamma}\boldsymbol{\gamma}', \tag{3.11}$$

respectively, where $\mathbb{V}\left(G_t \mid \boldsymbol{\Phi}_{t-1}\right) = \mathbb{E}\left[G_t^2 \mid \boldsymbol{\Phi}_{t-1}\right] - \left(\mathbb{E}\left[G_t \mid \boldsymbol{\Phi}_{t-1}\right]\right)^2$. Analogously, $\boldsymbol{\Gamma}$ is a correlation matrix only conditionally on the realization of the mixing process. For this reason, we call $\boldsymbol{\Gamma}$ the dependency matrix.

While $\boldsymbol{\Gamma}$ in (4.7) is not dynamic, the conditional correlation matrix of $\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1}$ is time-varying when $\boldsymbol{\gamma} \neq \mathbf{0}$ and $\mathbb{E}\left[G_t \mid \boldsymbol{\Phi}_{t-1}\right] \neq \mathbb{V}\left(G_t \mid \boldsymbol{\Phi}_{t-1}\right)$. If $\boldsymbol{\gamma} = \mathbf{0}$ or $\mathbb{E}\left[G_t \mid \boldsymbol{\Phi}_{t-1}\right] = \mathbb{V}\left(G_t \mid \boldsymbol{\Phi}_{t-1}\right)$ (e.g., in the MALap distribution below), then $\mathrm{Corr}\left(\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1}\right) = \boldsymbol{\Gamma}$, or $\mathrm{Corr}\left(\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1}\right) = \boldsymbol{\Gamma} + \boldsymbol{\gamma}\boldsymbol{\gamma}'$, respectively, and the dynamics in the parameters of $G_t \mid \boldsymbol{\Phi}_{t-1}$ in (4.4) influence only the variances.

All the conditional moments implied by the model (including the limiting cases of the mixing law) are available in Scott et al. (2011). The co-skewness and co-kurtosis matrices are also tractable; see the impressive expressions given in Mencía and Sentana (2009).

Finally, the unconditional mean and covariance of $\mathbf{Y}_t$ can be expressed in terms of the unconditional mean of G_t and unconditional covariance function of $\mathbf{Y}_t \mid G_t$, respectively as $\mathbb{E}[\mathbf{Y}_t] = \boldsymbol{\mu} + \mathbb{E}[G_t]\boldsymbol{\gamma}$ and $\mathrm{Cov}(\mathbf{Y}_t) = \mathbb{E}[\mathrm{Cov}(\mathbf{Y}_t \mid G_t)] + \mathbb{V}(G_t)\boldsymbol{\gamma}\boldsymbol{\gamma}'$.

(iii) From (4.16) it follows that the vector of conditional volatilities, defined as the square root of the conditional variances, and denoted by $\mathrm{vol}_{t|t-1}$, is given by

$$\mathrm{vol}_{t|t-1} = \sqrt{(\mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}]\,\mathbf{S}_t^2 + \mathbb{V}(G_t \mid \boldsymbol{\Phi}_{t-1})\,\boldsymbol{\gamma}^2)}. \tag{3.12}$$

The asymmetric effect in volatility states that the effects of positive returns on volatility are different from those of negative returns of a similar magnitude. The leverage effect refers to the negative correlation between the current return and future volatility. Therefore, leverage implies asymmetry, but not all asymmetric effects display leverage (see Asai and McAleer 2011; and the reference therein). In the GJR-GARCH model, both positive and negative returns increase future volatility, but the positive returns do so by less than the negative returns. Recalling that $\mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}] > 0$ and $\mathbb{V}(G_t \mid \boldsymbol{\Phi}_{t-1}) > 0$, it follows, from (3.12), that the only possibility to introduce leverage into the model is to use the dynamics for $s_{k,t}$ which support the leverage. In particular, neither the common market factor G_t, nor its dynamics, can be the source of leverage in the model.

(iv) As the volatility shock of the SV component is univariate, it influences (in a multiplicative way) each of the asset-specific conditional volatilities via (4.1b). Moreover, it drives the dynamics of higher conditional moments (e.g., skewness and kurtosis) and co-moments of the returns. One could argue that modeling volatility shocks with a univariate process is not sufficient because the reaction of the asset-specific volatilities to the common shock should vary across assets and through time. Nevertheless, because of the asset-specific conditional asymmetry coefficient γ_k, the impact of the SV component on each volatility is not equal across assets and its conditional expected value varies over time. (See Section 3.6.1 for an empirical demonstration of this.)

(v) There is a minor identification problem which needs to be addressed. The same MGHyp distribution arises from the parameter constellation $(\lambda_t, \chi_t/c, c\psi_t, \boldsymbol{\mu}, c\mathbf{H}_t, c\boldsymbol{\gamma})$ for any $c > 0$. One way to deal with this is to constrain the determinant of $\mathbf{H}_t$ to some particular value when fitting (see, e.g., McNeil et al., 2005). Alternatively, Protassov (2004) sets χ to a constant in his EM algorithm. We follow this latter approach and fix one of the GIG parameters prior to the estimation, as it is numerically simpler to implement in the iid (non-GARCH) setting, and, more crucially, because in our general model, the conditional dispersion matrix is time dependent. Moreover, this reduces the number of necessary equations to identify the parameters of $G_t \mid \boldsymbol{\Phi}_{t-1}$.

(vi) In order to estimate the dynamics in (4.4), the starting values of $\zeta_{r,t}$ and $\mathbb{E}\left[G_t^r \mid \mathbf{\Phi}_{t-1}\right]$ have to be set. In our empirical analysis, at each iteration in the estimation, we set them equal to the unconditional expected values, i.e., $\zeta_{k,0} = 0$ for all k and $\mathbb{E}\left[G_1^r \mid \mathbf{\Phi}_0\right] = c_r/\left(1-\rho_r\right)$, respectively. In addition, we assume that the roots of the polynomials $(1-\rho_r L)$, where L is the lag operator associated with (4.4), lie outside the unit circle (i.e., modulus $|\rho_r| < 1$), so that the unconditional expected value of G_t^r exists.

(vii) For notational convenience later, we collect the parameters of the model into three blocks (process, distribution, and correlation) as follows:

$$\boldsymbol{\theta}_P = (\boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}), \ \boldsymbol{\theta}_D = (\mathbf{c}, \boldsymbol{\rho}) \text{ and } \boldsymbol{\theta}_C = \mathbf{\Gamma}, \tag{3.13}$$

where $\boldsymbol{\omega}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}$ are K-dimensional vectors of GJR-GARCH$(1,1)$ parameters from (4.9) (the GARCH case (4.8) corresponds to $\eta_k = 0$); and $\mathbf{c}$ and $\boldsymbol{\rho}$ denote vectors of c_r and ρ_r parameters, respectively, though note that, in the case of constant G_t parameters, $\boldsymbol{\theta}_D$ reduces to just the set (λ, χ, ψ), of which in our three special cases, two are fixed.

3.3 Estimation

The explicit form of the MGHyp density and the structure of (4.1) implies that, if $\boldsymbol{\gamma} = \mathbf{0}$ (necessary to make the scale shock $\boldsymbol{\varepsilon}_t$ independent of the unobserved realizations of G_t), then the estimation of the parameters in the model can be performed by direct maximization of the corresponding likelihood function.[1] In particular, with $\boldsymbol{\theta}_P^S = \boldsymbol{\theta}_P \setminus \boldsymbol{\gamma}$, vector $\mathbf{Y}_t$ is assumed to have an MGHyp distribution with density $f_{\mathbf{Y}_t}\left(\mathbf{y}; \boldsymbol{\theta}_P^S, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C\right)$ given by

$$\int_0^\infty \frac{1}{(2\pi)^{K/2}\left|\mathbf{H}_t\right|^{1/2} g^{K/2}} \exp\left\{-\frac{(\mathbf{y}-\boldsymbol{\mu})' \mathbf{H}_t^{-1}(\mathbf{y}-\boldsymbol{\mu})}{2g}\right\} \times f_G(g; \boldsymbol{\theta}_D) \mathrm{d}g,$$

where $f_G(\cdot)$ is the density of the unobserved mixing random variable G given in (4.2); $\mathbf{H}_t$ admits the decomposition (4.7) with constant conditional correlation; and the dynamics of the scale terms, $s_{k,t}$, are from (4.8) or (4.9). One can then evaluate this integral to yield the explicit expression

$$f_{\mathbf{Y}_t}\left(\mathbf{y}; \boldsymbol{\theta}_P^S, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C\right) = C_t d_t^{-K/2+\lambda} K_{\lambda-K/2}(d_t), \tag{3.14}$$

for $d_t = \sqrt{\left(\chi + (\mathbf{y}-\boldsymbol{\mu})' \mathbf{H}_t^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)\psi}$ and the normalizing constant

$$C_t = \frac{\left(\sqrt{\chi\psi}\right)^{-\lambda} \psi^{K/2}}{(2\pi)^{K/2}\left|\mathbf{H}_t\right|^{1/2} K_\lambda\left(\sqrt{\chi\psi}\right)}.$$

Direct maximization of the likelihood function of $\mathbf{Y}$, denoted $L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P^S, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C\right)$, requires estimation of all the parameters in one step. This approach works for K very small, but

[1] The following estimation algorithm maximizes the conditional likelihood function, to ease the notation we omit the conditioning on $\mathbf{\Phi}_{t-1}$ across the derivation.

becomes problematic as the number of assets increases, due to the quadratic increase in the number of parameters associated with the dispersion matrix, and the linear increase due to $\boldsymbol{\mu}$ and the univariate GARCH parameters. In the case with $\boldsymbol{\gamma} \neq \mathbf{0}$ and/or for large K, direct maximization is no longer feasible. However, estimation can be conducted via an Expectation-Conditional Maximization Either (ECME) algorithm from Liu and Rubin (1994). The standard ECME algorithm maximizes the likelihood function by an iterative procedure. It is a fixed-point algorithm and consists of the E-step, in which the realizations of the unobserved mixing variables $\{G_t\}_{t=1,\ldots,T}$ are imputed; and the CM-steps, in which all the parameters are updated by maximizing either the unconditional likelihood function $L_{\mathbf{Y}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C)$, or the conditional one, $L_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C)$. This is now detailed.

The complete log-likelihood function is

$$\log L_{\mathbf{Y},\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C) = \log L_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C) + \log L_{\mathbf{G}}(\boldsymbol{\theta}_D), \tag{3.15}$$

where, based on the observed values $\mathbf{Y}_t = \mathbf{y}_t$ and conditional on $G_t = g_t = \widehat{G}_t$ (where the hatted notation indicates that G_t is estimated, and not observed), $t = 1, 2, \ldots, T$,

$$\begin{aligned}\log L_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C) = -\frac{1}{2}\sum_{t=1}^{T}\Big[&K\log(2\pi) + \log|g_t\mathbf{S}_t\boldsymbol{\Gamma}\mathbf{S}_t| \\ &+ g_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma}g_t)'\mathbf{S}_t^{-1}\boldsymbol{\Gamma}^{-1}\mathbf{S}_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma}g_t)\Big],\end{aligned} \tag{3.16}$$

and $L_{\mathbf{G}}(\boldsymbol{\theta}_D)$ denotes the likelihood function of $(G_t \mid \boldsymbol{\Phi}_{t-1}) \sim \mathrm{GIG}(\lambda_t, \chi_t, \psi_t)$, $t = 1, 2, \ldots, T$. In a similar fashion to Bollerslev (1990), Engle and Sheppard (2002), and Pelletier (2006), we split (3.16) into a sum of two terms,

$$\log L_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C) = \log L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P) + \log L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C), \tag{3.17}$$

where $L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P)$ is the mean-volatility term given by

$$\begin{aligned}\log L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P) = -\frac{1}{2}\sum_{t=1}^{T}\Big[&K\log(2\pi) + \log|\mathbf{S}_t|^2 \\ &+ g_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma}g_t)'\mathbf{S}_t^{-1}\mathbf{S}_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma}g_t) + \log g_t\Big],\end{aligned} \tag{3.18}$$

and $L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C)$ is the correlation term given by

$$\log L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log|\boldsymbol{\Gamma}| + \boldsymbol{e}_t'\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_t - \boldsymbol{e}_t'\boldsymbol{e}_t\right], \tag{3.19}$$

where $\boldsymbol{e}_t = g_t^{-1/2}\mathbf{S}_t^{-1}\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma}g_t$ from (4.1).

Owing to the mixture structure of the MGHyp, $L_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C)$ is a multivariate Gaussian likelihood with a GARCH-type structure for the scales and a given conditional correlation model. As such, maximization of $L_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C)$ can be done in two steps. First, with the correlation structure ignored, the GARCH parameters in (4.8) or (4.9) are estimated for

each of the K assets separately (and concurrently with parallel computing) by maximizing $L_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P \mid \boldsymbol{\theta}_C = \mathbf{I}_K)$, and, in the second step, the correlation parameters $\boldsymbol{\theta}_C$ are estimated from the first-step de-volatilized residuals. This idea is not new, and is similar to the Gaussian setup in Bollerslev (1990), Engle (2002), and Pelletier (2006). Given the $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_C$ estimates, denoted by $\widehat{\boldsymbol{\theta}}_P$ and $\widehat{\boldsymbol{\theta}}_C$, the mixing process parameters $\boldsymbol{\theta}_D$ are estimated by maximizing $L_{\mathbf{Y}}\big(\boldsymbol{\theta}_D \mid \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_C\big)$. Given all the estimates, we proceed with the next E-step update (see (4.21) below) of the unobserved mixing random variable $\mathbf{G}$, and continue to iterate until convergence.

Observe that $L_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C)$ reduces to the mean-volatility component, $L^{\text{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P)$, if and only if we assume zero correlations. As such, $L^{\text{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P)$ corresponds to the likelihood of $\mathbf{Y}$ conditional on $\mathbf{G}$ for the model under zero correlations. Based on this decomposition, estimates of $\boldsymbol{\theta}_P$, $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_D$ can be obtained by the following ECME algorithm.

E-step: Calculate $\mathbb{E}\big[\log L_{\mathbf{Y},\mathbf{G}} \mid \mathbf{Y}; \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_D, \widehat{\boldsymbol{\theta}}_C\big]$.

The log-likelihood function (4.19) is linear with respect to g_t and g_t^{-1} ($\log g_t^{-1}$ can be ignored without influencing the first order conditions). Hence the E-step involves replacing unobserved realizations of G_t and G_t^{-1} in (4.18) by the imputed values, $\widehat{G}_t$. Calculation shows that (see, e.g., Paolella, 2013, Eq. 35)

$$\Big(G_t \mid \boldsymbol{\Phi}_t; \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_C, \widehat{\boldsymbol{\theta}}_D\Big) \sim \text{GIG}\left(\lambda_t^*, \chi_t^*, \psi_t^*\right), \tag{3.20}$$

where

$$\lambda_t^* = \lambda_t - K/2,\ \chi_t^* = \chi_t + (\mathbf{y}_t - \widehat{\boldsymbol{\mu}})' \mathbf{S}_t^{-1} \widehat{\boldsymbol{\Gamma}}^{-1} \mathbf{S}_t^{-1} (\mathbf{y}_t - \widehat{\boldsymbol{\mu}}) \text{ and } \psi_t^* = \psi_t + \widehat{\boldsymbol{\gamma}}' \mathbf{S}_t^{-1} \widehat{\boldsymbol{\Gamma}}^{-1} \mathbf{S}_t^{-1} \widehat{\boldsymbol{\gamma}}.$$

The latent values of g_t and g_t^{-1} are then updated by their conditional expectations from (4.21), using the expression for the moments of the GIG random variable given in (4.6).

CM1-step: Update $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_C$.

(P) Update $\boldsymbol{\theta}_P$ by computing

$$\arg\max_{\boldsymbol{\theta}_P} \log L^{\text{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P), \tag{3.21}$$

where $L^{\text{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P)$ is a Gaussian likelihood with zero correlation, so we can estimate the parameters of each asset, $(\mu_k, \gamma_k, \omega_k, \alpha_k, \beta_k)$, separately by maximizing the corresponding likelihood function.

(C) Update $\boldsymbol{\theta}_C$ by computing the usual empirical correlation estimator (the MLE under normality) of the de-volatilized residuals $\widehat{G}_t^{-1/2} \mathbf{S}_t^{-1} \widehat{\boldsymbol{\varepsilon}}_t$, where $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\gamma}} \widehat{G}_t$, $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\gamma}}$ are obtained in part (P) directly above, and $\widehat{G}_t$ is obtained in the E-step.

CM2-step: Given the CM1-step estimates of $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_C$, obtain new estimates of $\boldsymbol{\theta}_D$ by maximizing the incomplete data log-likelihood function, i.e., compute

$$\arg\max_{\boldsymbol{\theta}_D} \log L_{\mathbf{Y}}\left(\boldsymbol{\theta}_D \mid \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_C\right). \tag{3.22}$$

Iterate the above steps until convergence.

Remarks:

(i) In contrast to the Gaussian distribution, setting $\mathbf{\Gamma}$ equal to the identity matrix does not imply independence when assuming MGHyp returns, due to the dependence induced by the GIG mixing variable G_t. But via the application of the ECME algorithm, which conditions on the realizations of G_t, we can estimate the parameters of the GARCH equations separately for each asset. This is the key to fast, simple, joint likelihood-based estimation of a multivariate non-normal model.

(ii) In the ECME algorithm, the E-step computation of the conditional expectations of G_t and G_t^{-1} involves computation of ratios of Bessel functions (4.3) for different arguments. In case of large arguments (χ_t^* and ψ_t^* in (4.21) can get very large because they involve a quadratic form of the inverse of the covariance matrix), numerical computation is subject to rounding error which affects estimates of all the parameters. We propose a method which increases numerical accuracy such that estimation for all data windows used in our empirical application was successful. It is given in Appendix 3.8.4.

(iii) The method which increases the accuracy of the Bessel function computation given in Appendix 3.8.4 is also used in the CM2-step for evaluating $\log L_{\mathbf{Y}}$ for different $\boldsymbol{\theta}_D$ values in (4.24).

(iv) Here we state the starting values used for the estimation procedure. Those of the asymmetry parameter γ_k and the GJR-GARCH parameter η_k are taken to be zero, and the remaining ones in $\boldsymbol{\theta}_P$, $(\mu_k, \omega_k, \alpha_k, \beta_k)$, to be those values obtained from the normal-based GARCH estimates, using the method in Paolella and Polak (2013a) to avoid inferior local likelihood maxima. For $\boldsymbol{\theta}_C$, we use the empirical correlation matrix computed from the normal-based GARCH residuals. For $\boldsymbol{\theta}_D = (\lambda, \chi, \psi)$, we have confirmed that the likelihood is such that optimization is rather robust to the choice of starting values. For the special case of the multivariate Laplace distribution (MALap) used in the empirical section below, we use $\lambda = 2$ as the starting value. For the hybrid GARCH-SV model we set, $c = 0.1$ and $\rho = 0.8$. In the estimation with a rolling window, we use also as starting values the previous window estimates; and take the final estimates to be those with the higher likelihood value.

(v) If the starting values are sufficiently close to those which maximize the likelihood, or if the likelihood function is unimodal in the parameter space, and the maximum is not on the boundary of the parameter space, then monotonicity in the likelihood values of the consecutive ECME estimates guarantees their convergence to the corresponding maximum likelihood parameters; see, e.g., McLachlan and Krishnan (2008). As such, under further standard regularity conditions on the likelihood, consistent and asymptotically normal point estimates are obtained.

(vi) For $\boldsymbol{\gamma} = 0$ and K small, we confirmed the accuracy of the proposed EM algorithm by comparing the EM estimates with the results based on the direct likelihood maximization of (3.14).

3.4 Special Cases of the MGHyp

Fixing some of the MGHyp distribution parameters in a judiciously chosen way can result in a distribution which is not only virtually as capable of modeling the features of financial data as the full general MGHyp case, but can even result in superior results in the same way that parameter shrinkage results in lower mean squared error. To this end, we use three special cases of the MGHyp distribution in which: (i) G_t is gamma distributed with shape parameter $\lambda > 0$ and unit scale parameter (the multivariate asymmetric Laplace, MALap, model); (ii) the λ parameter is fixed at $-1/2$ (the multivariate normal inverse Gaussian, MNIG, model); and (iii) G_t is inverse gamma distributed with the scale and the shape parameters both equal to $v/2$ (equivalently G_t is GIG with $\lambda = -0.5v$, $\chi = v$ and $\psi = 0$) where v, $v > 0$, is the parameter being estimated (the multivariate asymmetric t-distribution, MAt, model).

As reported by Protassov (2004), and also confirmed by our studies, one or more of the MGHyp shape parameters can have a relatively flat likelihood (already after fixing χ or ψ for identification purposes), implying possible numeric problems when maximizing the likelihood. The three special cases do not share the flat likelihood problem of the fully general MGHyp distribution, and so are considerably faster and numerically more reliable to estimate, but still retain the flexibility required for modeling asset returns by allowing for individual asset asymmetry parameters and also higher kurtosis than the normal distribution. Section 3.4.1 details the derivation and properties of the MALap density, while Sections 3.4.2 and 3.4.3 examine the MNIG and MAt distributions, respectively.

3.4.1 Multivariate Laplace Distribution

The MALap density can be derived by evaluating the integral

$$f_{\mathbf{Y}}(\mathbf{y}; \lambda, \boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\gamma}) = \int f_{\mathbf{Y}|G}(\mathbf{y} \mid g; \boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\gamma}) f_G(g; \lambda) \mathrm{d}g;$$

where $G \sim \mathrm{Gam}(\lambda, 1)$; or from the MGHyp density (3.14), with $\lambda > 0$, $\chi = 0$ and $\psi = 2$, by use of the limiting relation $K_\lambda(\sqrt{(2\chi)}) \simeq \Gamma(\lambda) 2^{\lambda-1} (\sqrt{2\chi})^{-\lambda}$ for $\chi \downarrow 0$, $\lambda > 0$ (Paolella, 2007, Eq. 9.6), where Γ is the gamma function. Both approaches result in $f_{\mathbf{Y}_t}(\mathbf{y}; \lambda, \boldsymbol{\mu}, \mathbf{H}_t, \boldsymbol{\gamma})$ given by

$$\frac{2 \exp\left\{(\mathbf{y} - \boldsymbol{\mu})' \mathbf{H}_t^{-1} \boldsymbol{\gamma}\right\}}{(2\pi)^{K/2} \Gamma(\lambda) |\mathbf{H}_t|^{1/2}} \left(\frac{m_t}{2 + \boldsymbol{\gamma}' \mathbf{H}_t^{-1} \boldsymbol{\gamma}} \right)^{\lambda/2 - K/4} K_{\lambda - K/2} \left(\sqrt{m_t \left(2 + \boldsymbol{\gamma}' \mathbf{H}_t^{-1} \boldsymbol{\gamma}\right)} \right), \qquad (3.23)$$

where $\mathbf{H}_t$ is given by (4.7) and $m_t = (\mathbf{y} - \boldsymbol{\mu})' \mathbf{H}_t^{-1} (\mathbf{y} - \boldsymbol{\mu})$. It generalizes the distribution proposed and used in Paolella (2013), which is (3.23) but without the asymmetry term $\boldsymbol{\gamma}$. It turns out that our discovery is not new: a symmetric and univariate version of this distribution was

introduced by McKay (1932); it was extended to the multivariate case, called variance-gamma and used in finance by Madan and Seneta (1990); it was also mentioned in Kotz et al. (2000, 2001) and Kozubowski and Podgórski (2001) as a generalized Laplace, but they concentrated on further special cases of it; for the univariate asymmetric version of it, see Paolella (2007, Ch. 9).

The MALap allows for flexible modeling of power and exponential factors of the tail of the distribution. Importantly, if $(\mathbf{Y}_t \mid \mathbf{\Phi}_{t-1}) \sim \text{MGHyp}(\boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{H}_t, \lambda_t, \chi_t, \psi_t)$, then the distribution of the portfolio return $P_t = \mathbf{w}'\mathbf{Y}_t$, with a vector of portfolio weights $\mathbf{w} \in \mathbb{R}^K \setminus \mathbf{0}$, is given by $(P_t \mid \mathbf{\Phi}_{t-1}) \sim \text{GHyp}(\mathbf{w}'\boldsymbol{\mu}, \mathbf{w}'\boldsymbol{\gamma}, \mathbf{w}'\mathbf{H}_t\mathbf{w}, \lambda_t, \chi_t, \psi_t)$. The limiting behavior of the GHyp density (see Barndorff-Nielsen 1997; Prause 1999, eq. 1.19), when $x \to \pm\infty$, is

$$f_{P_t}\left(x; \mu_{\mathbf{w}}, \gamma_{\mathbf{w}}, \sigma^2_{t,\mathbf{w}}, \lambda_t, \chi_t, \psi_t\right) \propto |x|^{\lambda_t - 1} \exp\left(-\sqrt{\frac{\psi_t + \gamma_{\mathbf{w}}/\sigma^2_{t,\mathbf{w}}}{\sigma^2_{t,\mathbf{w}}}}\,|x| + \frac{\gamma_{\mathbf{w}}}{\sigma^2_{t,\mathbf{w}}} x\right), \tag{3.24}$$

where $\mu_{\mathbf{w}} = \mathbf{w}'\boldsymbol{\mu}$, $\gamma_{\mathbf{w}} = \mathbf{w}'\boldsymbol{\gamma}$, and $\sigma^2_{t,\mathbf{w}} = \mathbf{w}'\mathbf{H}_t\mathbf{w}$. It is a product of the power-tail component $|x|^{\lambda_t - 1}$ and the exponential. The latter is described by the following three parameters: (i) the dispersion matrix, $\mathbf{H}_t$ (with GARCH effects and the correlation matrix $\mathbf{\Gamma}$); (ii) the asymmetry vector $\boldsymbol{\gamma}$; and (iii) the GIG distribution parameter, ψ_t ($\psi_t = 2$ in the MALap case). The power-tail factor is described by only one parameter, λ_t, which is estimated, so that the tail thickness of the data can be captured by the model. In the conditional setup as given in Section 4.2 above, the exponential-tail dynamics are well-described by the GARCH structure in $\mathbf{H}_t$ and, in the case of the MALap distribution, the hybrid GARCH-SV parametrization given in (4.4) also allows for possible power-tail dynamics.

3.4.2 Multivariate Normal Inverse Gaussian Distribution

The MNIG distribution arises from the MGHyp for $\lambda = -1/2$, with density

$$f_{\mathbf{Y}_t}\left(\mathbf{y}; \boldsymbol{\mu}, \mathbf{H}_t, \boldsymbol{\gamma}, \chi, \psi\right) = C_t d_t^{-(K+1)/2} K_{(K+1)/2}(d_t) \exp\left\{(\mathbf{y} - \boldsymbol{\mu})' \mathbf{H}_t^{-1} \boldsymbol{\gamma}\right\}, \tag{3.25}$$

for $d_t = \sqrt{\left(\chi + (\mathbf{y} - \boldsymbol{\mu})' \mathbf{H}_t^{-1} (\mathbf{y} - \boldsymbol{\mu})\right)\left(\psi + \boldsymbol{\gamma}'\mathbf{H}_t^{-1}\boldsymbol{\gamma}\right)}$ and the normalizing constant

$$C_t = \frac{(\chi/\psi)^{1/4} \left(\psi + \boldsymbol{\gamma}'\mathbf{H}_t^{-1}\boldsymbol{\gamma}\right)^{(K+1)/2}}{(2\pi)^{K/2} |\mathbf{H}_t|^{1/2} K_{1/2}\left(\sqrt{\chi\psi}\right)}.$$

This distribution has been used in the iid case (McNeil et al., 2005) and also for building a multivariate predictive distribution via univariate NIG-GARCH components (Broda and Paolella, 2009). In our model, for identification purposes (see Remark 4.2.(v)) we additionally fix $\psi = 1$. In our forecasting study below, we show that, while use of the MNIG is far better than use of the normal, the model using the MALap distribution and the hybrid dynamics results in better performance.

3.4.3 Multivariate t-Distribution

The MAt is a limiting case of the MGHyp, with $\lambda = -1/2v$, $\chi = v$ and $\psi = 0$, for some $v > 0$. Evaluating the limit of (3.14) as $\psi \to 0$ gives the density

$$f_{\mathbf{Y}_t}(\mathbf{y}; \boldsymbol{\mu}, \mathbf{H}_t, \boldsymbol{\gamma}, v) = C_t \frac{K_{(v+K)/2}\left(\sqrt{(v+m_t)\,\boldsymbol{\gamma}'\mathbf{H}_t^{-1}\boldsymbol{\gamma}}\ \exp\left((\mathbf{y}-\boldsymbol{\mu})'\,\mathbf{H}_t^{-1}\boldsymbol{\gamma}\right)\right)}{\left(\sqrt{(v+m_t)\,\boldsymbol{\gamma}'\mathbf{H}_t^{-1}\boldsymbol{\gamma}}\right)^{-(v+K)/2}\left(1+m_t/v\right)^{(v+K)/2}}, \tag{3.26}$$

where $m_t = (\mathbf{y}-\boldsymbol{\mu})'\,\mathbf{H}_t^{-1}\,(\mathbf{y}-\boldsymbol{\mu})$, and normalizing constant

$$C_t = \frac{2^{1-(v+K)/2}}{\Gamma(v/2)\,(\pi v)^{K/2}\,|\mathbf{H}_t|^{1/2}}.$$

This density reduces to the standard multivariate t with $v/2$ degrees of freedom, as $\boldsymbol{\gamma} \to \mathbf{0}$.

McNeil et al. (2005, Ch. 3) mention this distribution as having potential applications in finance, while Aas and Haff (2006b) work with the univariate version of (3.26), and show that it is the only subclass of the GHyp family that has the property of different asymptotic left and right tail behavior. The heavier tail has a power decay and the lighter tail is a product of a power and an exponential function. Aas and Haff (2006b) show that, in the univariate case and under an iid assumption, it leads to superior data fit than some other competitors, including the NIG distribution. Jondeau (2012) modifies this distribution and uses it to show the importance of asymmetry in the tail dependence of equity portfolios.

3.5 Option Pricing

We present a feasible technique which allows for multivariate option pricing in the framework of our model. Since the work of Duan (1995), GARCH models have become increasingly popular in option pricing. More recent literature includes Heston and Nandi (2000), who derive a nearly closed-form pricing formula under normal return innovations and the valuation assumption from Duan (1995); Christoffersen et al. (2006), who propose a model with inverse Gaussian innovations which allows for conditional skewness; Barone-Adesi et al. (2008), who use filtered historical simulation; Christoffersen et al. (2010), who develop a theoretical framework for option valuation under very general assumptions which allow for conditional heteroskedasticity and non-normality; and Rombouts and Stentoft (2011), who consider multivariate option pricing in a model with a finite-mixture-of-normal. Our model is also multivariate and, as it allows for all the primary stylized facts of asset returns, it is expected to be a good candidate for option pricing, given a feasible calibration algorithm.

The proposed algorithm combines the equivalent martingale measure (EMM) technique in the presence of a GARCH structure as in Christoffersen et al. (2010), with an iterative estimation of the model dynamics. Like in Barone-Adesi et al. (2008), it does not focus on the analytical form of the change of measure. Barone-Adesi et al. (2008) utilize a Monte

Carlo simulation based on QML estimates of model parameters under the historical measure and calibrate the EMM on the option prices. In contrast to their nonparametric method, our algorithm estimates the model parameters via maximizing the complete data likelihood function under the historical measure P, it changes the measure as if G_t were observed, and it iteratively replaces the missing information about G_t with conditional expectation, under the new measure, of G_t given $\mathbf{Y}_t$. Crucially for the COMFORT model, the proposed algorithm does not suffer from the curse of dimensionality, and so, it is applicable in a multivariate setup with a large number of underlying assets.

Denote by $\mathbf{X}_t = (X_{t,1}, X_{t,2}, \ldots, X_{t,K})'$ a vector of prices of assets $k = 1, 2, \ldots, K$ at time t. The price of an option contract at time t with maturity T and terminal payoff function $\varrho_\vartheta(\mathbf{X}_T)$, ϑ being the set of relevant parameters such as the strike price, can be computed as the following discounted expectation,

$$C_t(T, \varrho_\vartheta) = \exp(-(T-t)\, r)\, \mathbb{E}^P\left[\varrho_\vartheta(\mathbf{X}_T)\, \frac{\mathrm{d}Q^*}{\mathrm{d}P} \mid \mathbf{\Phi}_t\right], \tag{3.27}$$

where $\frac{\mathrm{d}Q^*}{\mathrm{d}P} \mid \mathbf{\Phi}_t$ is the change of measure such that the discounted stock price process under Q^* is a martingale with respect to the $\mathbf{\Phi}_t$ filtration. This filtration is defined in Section 4.2 and it is associated with the incomplete conditional density function $f_{\mathbf{Y}_t}$. For the purpose of option pricing, we define a complete information filtration,[1] associated with the complete information density $f_{\mathbf{Y}_t, G_t}$, by $\mathcal{F}_t = \sigma(\{G_1, \mathbf{Y}_1, G_2, \mathbf{Y}_2, \ldots, G_t, \mathbf{Y}_t\})$; and a second filtration, which includes information about the realization of G_{t+1}, $\mathcal{F}_t^{+G} = \sigma(\{G_1, \mathbf{Y}_1, G_2, \mathbf{Y}_2, \ldots, G_t, \mathbf{Y}_t, G_{t+1}\})$. Hence, $\mathbf{\Phi}_t \subseteq \mathcal{F}_t \subseteq \mathcal{F}_t^{+G}$ and, more generally, for any t, the following *chain property* for the latter two filtrations holds,

$$\mathcal{F}_1 \subseteq \mathcal{F}_1^{+G} \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_2^{+G} \subseteq \ldots \subseteq \mathcal{F}_t \subseteq \mathcal{F}_t^{+G}. \tag{3.28}$$

The mixture property of the MGHyp distribution implies that $\mathbf{Y}_t \mid \mathcal{F}_{t-1}^{+G} \sim \mathrm{N}(\boldsymbol{\mu} + \boldsymbol{\gamma} G_t, G_t \mathbf{H}_t)$, hence the standard theory for option pricing under normality applies. In particular, following Christoffersen et al. (2010), and Rombouts and Stentoft (2011), if we impose the exponential affine form on the Radon-Nikodym derivative, with respect to $\mathcal{F}_t^{+G}$, then, under the corresponding measure Q^{+G}, as detailed in Appendix 3.8.2, the dynamics of the returns remain Gaussian although with a shift in the mean.

Next, we define a change of measure, under $\mathcal{F}_t$, by

$$\frac{\mathrm{d}Q}{\mathrm{d}P} \mid \mathcal{F}_t = \mathbb{E}^P\left[\left(\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_t^{+G}\right) \mid \mathcal{F}_t\right]. \tag{3.29}$$

This transformation defines a Radon-Nikodym derivative under $\mathcal{F}_t$ and the resulting measure Q is an EMM, under $\mathcal{F}_t$. Note that, from (3.38) the information contained in $\mathcal{F}_t$ is sufficient

[1] Note an important difference that the complete information filtration does not imply directly that the market is complete. The discreteness of time in the model, the fact that asset prices can "jump" to infinitely many values from day to the next, the presence of time varying volatilities and only a finite number of instruments (options, underlying asset prices, risk-free bond) that are available in the market induce incompleteness.

to construct $\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_t^{+G}$. Therefore, the change of measure (3.29) is available explicitly and equal to (3.38). Under the measure Q

$$\mathbf{Y}_t \mid \mathcal{F}_{t-1}^{+G} \overset{Q}{\sim} \mathrm{N}\left((\boldsymbol{\mu} - \mathbf{r}) + \boldsymbol{\gamma} G_t + \frac{1}{2}\mathrm{diag}\left(\mathbf{S}_t^2\right) G_t,\ G_t \mathbf{H}_t\right), \tag{3.30}$$

where $\mathbf{r} = (r, \ldots, r)'$ is a $K \times 1$ vector of the risk free interest rates r.

Under no arbitrage conditions, the derivatives which are a function of the underling stock price process can be priced as the expected value, under EMM, of their future cash flows discounted using the risk free interest rate, as given in (3.27). So far we have derived the necessary tools for pricing the option under the condition that the realizations of the G_t sequence are observed. In particular, note that $\mathcal{F}_{t-1}^{+G} \nsubseteq \boldsymbol{\Phi}_t$ and $\mathcal{F}_{t-1} \nsubseteq \boldsymbol{\Phi}_t$, so a more general chain relation, than (3.28), does not hold, and one cannot rely on a simple two step procedure to obtain the EMM under $\boldsymbol{\Phi}_t$.

The algorithm in Appendix 3.8.3 mitigates this problem. Instead of analytically deriving the change of measure, it estimates the parameters of the model under the EMM Q^*. This method implicitly defines the change of measure, and it is feasible even in case of a large number of assets. Moreover, it is expected to reduce the estimation error (even if there would be a feasible *and* consistent two step approach) because it directly focuses on the Q^* measure parameters instead of doing this in two steps (first ML under historical measure P, and then calibration of the measure change).

Having obtained, the parameter estimates under the risk neutral measure Q^* the conditional distribution of the returns is given by

$$\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1} \overset{Q^*}{\sim} \mathrm{MGHyp}\left(\boldsymbol{\mu}^{Q^*}, \boldsymbol{\gamma}^{Q^*}, \mathbf{H}_t^{Q^*}, \boldsymbol{\theta}_D^{Q^*}\right), \tag{3.31}$$

where $\mathbf{H}_t^{Q^*} = \mathbf{S}_t^{Q^*} \boldsymbol{\Gamma}^{Q^*} \mathbf{S}_t^{Q^*}$ is the dispersion matrix under Q^*, with the scale term dynamics, in a diagonal matrix $\mathbf{S}_t^{Q^*}$, with the $\boldsymbol{\theta}_P$ parameters under Q^*.

Given the distribution of the returns under the risk neutral measure, one can price various options by using (3.27), and Monte Carlo simulation.

3.6 Empirical Application

To demonstrate the applicability and competitiveness of the model, we use the data set consisting of the $2,767$ daily returns of $K = 30$ components of the Dow Jones Industrial Index (DJ-30) from January 2nd, 2001, to December 30th, 2011 (based on the DJ-30 composition as of June 8th, 2009). Returns for each asset are computed as continuously compounded percentage returns, given by $y_{k,t} = 100 \log\left(p_{k,t}/p_{k,t-1}\right)$, where $p_{k,t}$ is the price of asset k at time t.

We first compare the in-sample fit of the usual, multivariate normal CCC (MN-CCC) model to the new MALap-CCC model, and show that the latter provides a much better fit to the tails of the return distribution. Next, we discuss the impact of the common market factor on the conditional volatilities. We conclude with the implications of the hybrid GARCH-SV extensions of our model.

We then compare the forecasting performance across different models. Summarizing the results, the MAt-CCC models and MALap-CCC hybrid GARCH(1, 1)-SV model deliver the best density predictions among all considered models. In particular, we find (i) a large improvement moving from the normal (even with dynamic correlations) to any of the distributions discussed in Section (3.4); and (ii) that the introduction of the dynamics in the G_t parameters (hybrid GARCH-SV model) further improves the forecasting performance. Both of these improvements are statistically significant. The overall best performing model is MALap-CCC hybrid GARCH(1, 1)-SV.

3.6.1 In-Sample Performance

We estimate the MN-CCC and the MALap-CCC models for the whole data sample and compare the in-sample fit by inspecting the Q-Q-plots (the sample quantiles of the standardized residuals versus the theoretical quantiles from a normal distribution) of the resulting estimated standardized residuals given by

$$\widehat{\mathbf{H}}_t^{-1/2}\big(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}}\big) \quad \text{and} \quad \widehat{G}_t^{-1/2}\widehat{\mathbf{H}}_t^{-1/2}\big(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\gamma}}\widehat{G}_t\big), \tag{3.32}$$

respectively; where $\widehat{G}_t$ are the imputed values of G_t returned from the ECME algorithm, $\widehat{\mathbf{H}}_t$ are fitted conditional dispersion matrices, and other hatted entries denote parameter estimates. Observe that both sets of residuals in (3.32), in particular the latter, are assumed to be Gaussian under each of the assumed models. In Figures 3.1 and 3.2 we provide the Q-Q plots of the residuals of three (Gaussian, MALap, and MAt based) competitive models, for JPMorgan Chase & Co, Bank of America, American Express, and Microsoft Corp, based on the entire sample of $T = 2,767$ observations. From Figure 3.1, it is apparent that the MALap-CCC Hybrid GARCH(1, 1)-SV model provides a markedly better (albeit not perfect) fit for the tail probabilities than the MN-CCC GARCH(1, 1) model. In Figure 3.2 we compare the fit of the MALap-CCC Hybrid GARCH(1, 1)-SV model and the MAt-CCC GARCH(1, 1) model. The latter model in the univariate case, usually called t-GARCH, is well known for providing an excellent model fit. Here, from Figure 3.2, we see that the results are comparable to those of the MALap-CCC Hybrid GARCH(1, 1)-SV model, though for very extreme events, the MAt performs slightly better.

Now consider the filtered G_t sequence. Figure 3.3 illustrates its impact on one of the assets, Merck & Co. The top panel gives the returns, while the second panel shows the filtered $\widehat{G}_t$ values from the ECME algorithm computed from (4.21). In the third panel, the scale-term, $s_{k,t}$, for the same asset, implied by the estimates of the GARCH(1, 1) dynamics from (4.8), are plotted over time. The panel in the last row combines the above factors and plots the $Y_{k;t}$ volatilities, as defined in (3.12) and based on the parameter estimates. A very negative spike in the second quarter of the data is synchronic with a large spike in the $\widehat{G}_t$ sequence in the second panel, which corresponds to the spike in the volatility in the last panel (especially when compared with the scale-term dynamics from the third panel). This illustrates the role of the common market factor as a stochastic latent filter. Interestingly, in the periods of high volatility

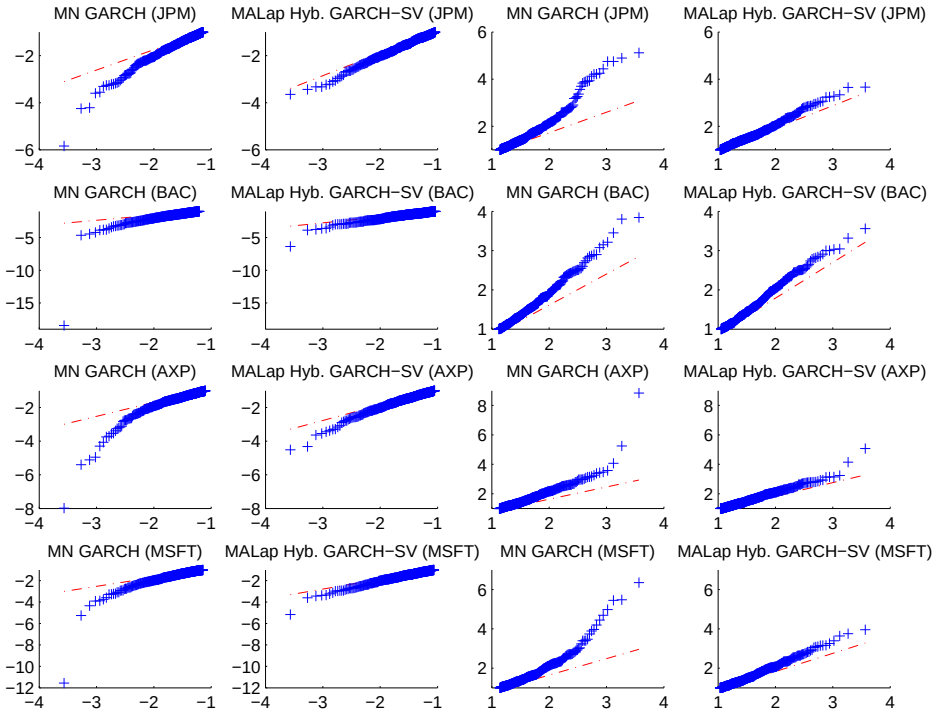

Figure 3.1: Tails of the quantile plots of the conditional distribution of innovations based on the $2,767$ observations. **Rows:** From top to bottom JPMorgan Chase & Co. (JPM); Bank of America (BAC); American Express (AXP); Microsoft Corp. (MSFT). **First column:** The left tail of the MN-CCC GARCH$(1,1)$ model. **Second column:** The left tail of the MALap-CCC Hybrid GARCH$(1,1)$-SV model. **Third column:** The right tail of the MN-CCC GARCH$(1,1)$ model. **Fourth column:** The right tail of the MALap-CCC Hybrid GARCH$(1,1)$-SV model.

(e.g., around the crisis of 2008) there are no strong market shocks. Although volatilities are very high, their magnitude is adequately accounted for by the GARCH$(1,1)$ dynamics, and the G_t factor is instead responsible for sharper volatility moves.

The effect of G_t across assets is not equal. From (4.16), each asset volatility is a sum of two terms. The first term is a product of the scale-term, $s_{k,t}$, and the conditional expected value of the common market factor, so the impact of the G_t term on each asset volatility depends on the level of the corresponding scale-term. The second term is a product of the conditional variance of G_t and the square of the asymmetry coefficient in the vector $\boldsymbol{\gamma}$. Hence, the impact of G_t on volatilities differs across assets. Figure 3.4 illustrates this fact. It is a multivariate analogue of Figure 3.3 and explains the contribution of the G_t factor in the conditional volatilities from the MALap-CCC hybrid GARCH$(1,1)$-SV model. Clearly, the G_t spikes have different impacts on volatilities of different assets.

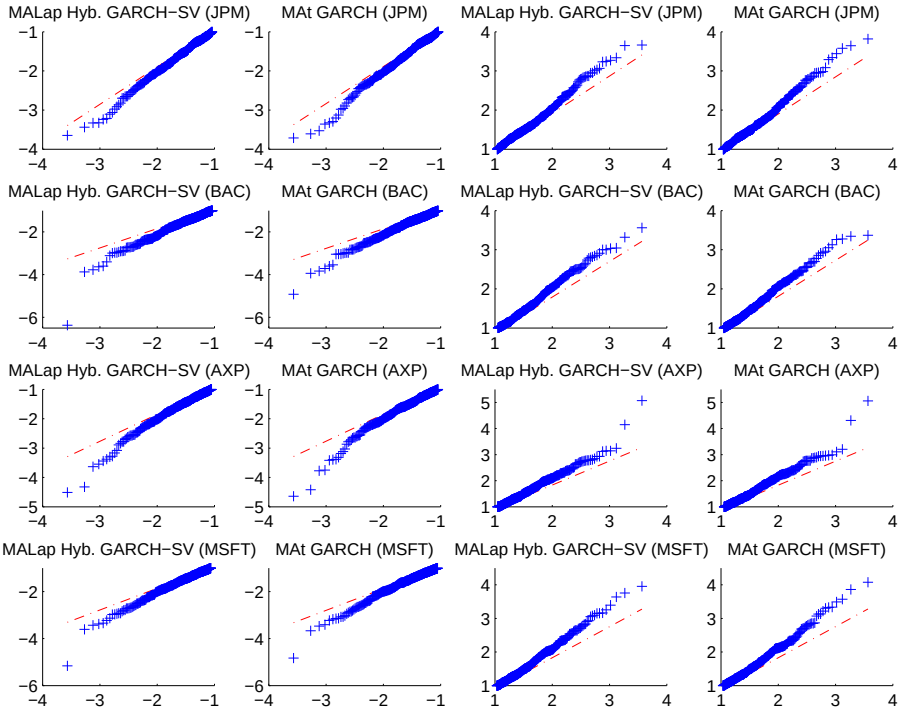

Figure 3.2: Tails of the quantile plots of the conditional distribution of innovations based on the $2,767$ observations, with rows analogous to Figure 3.1. **First column:** The left tail of the MALap-CCC Hybrid GARCH$(1,1)$-SV model. **Second column:** The left tail of the MAt-CCC GARCH$(1,1)$ model. **Third column:** The right tail of the MALap-CCC Hybrid GARCH$(1,1)$-SV model. **Fourth column:** The right tail of the MAt-CCC GARCH$(1,1)$ model.

We now discuss the consequences of the hybrid GARCH$(1,1)$-SV extension. The sequence of unobserved mixing random variables G_t implies the non-normality of the model and, in general, cannot be predicted. The role of the SV extension is to filter, through the dynamics in (4.4), a possible persistence in G_t. We model only the dynamics in the parameters of G_t, and not the dynamics of G_t itself. The consequence of this is that we need to distinguish between $\mathbb{E}\left[G_t \mid \mathbf{\Phi}_{t-1}\right]$ and $\mathbb{E}\left[G_t \mid \mathbf{\Phi}_t\right]$. The former are either constant over time (when G_t are iid) or, with $G_t \mid \mathbf{\Phi}_{t-1}$ having time-varying parameters. The latter, $\mathbb{E}\left[G_t \mid \mathbf{\Phi}_t\right]$, are filtered from the E-step update of the ECME algorithm. They condition on the observed data up to *and including* time t and, obviously, cannot be used for prediction, but instead they serve as a natural benchmark to judge the in-sample-fit of $\mathbb{E}\left[G_t \mid \mathbf{\Phi}_{t-1}\right]$.

In the first two panels of Figure 3.5, we compare $\mathbb{E}\left[G_t \mid \mathbf{\Phi}_t\right]$ and $\mathbb{E}\left[G_t \mid \mathbf{\Phi}_{t-1}\right]$ based on the MALap-CCC models. The case with iid G_t is given in the first panel. In the iid case, $\mathbb{E}\left[G_t \mid \mathbf{\Phi}_{t-1}\right] = \mathbb{E}\left[G_t\right]$, and we see that they result in a relatively poor fit. The second panel is

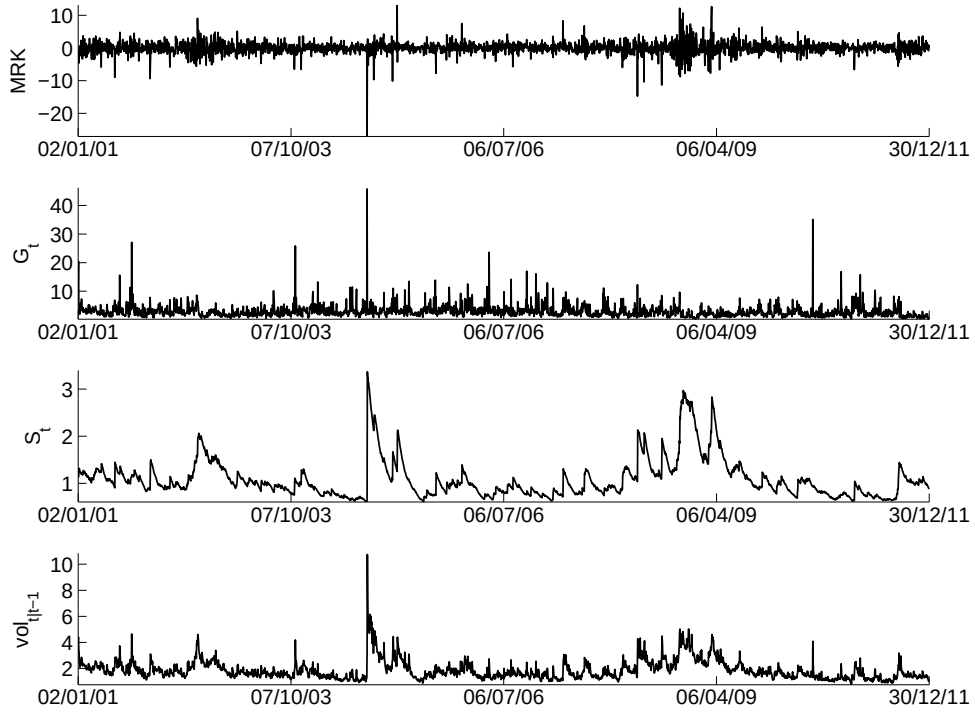

Figure 3.3: The impact of the common market factor on one of the assets (Merck & Co). **First row:** Returns $Y_{k,t}$ of Merck & Co. **Second row:** Values of the filtered common market factor $\widehat{G}_t$ from the ECME algorithm. **Third row:** The scale-term, $s_{k,t}$, for the same asset, implied by the estimates of the GARCH$(1,1)$ model. **Fourth row:** The conditional volatility of $Y_{k,t}$, computed as the square root of the kth element on the diagonal of matrix (4.16) and based on the parameter estimates.

for the hybrid extension of the model, where the dynamics of the $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$ are described by (4.4). The latter model clearly results in better fit of the common market factor. The $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$ estimates match the filtered values and even the largest spikes (which could be interpreted as describing highly unexpected news) are well-accommodated.

The last two panels in Figure 3.5 compare the resulting conditional volatilities from the two models. Again, the conditional volatilities from the hybrid extension (computed from (4.16) with use of $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$) lie much closer to the filtered values (computed from (4.16) with use of $\mathbb{E}[G_t \mid \mathbf{\Phi}_t]$).

What is common to all assets is that the G_t factor explains a large fraction of the volatility. Based on the whole sample estimates, Figure 3.6 displays the correlations between $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$ and the conditional volatilities of the assets filtered from the ECME algorithm. Remarkably, for 26 out of 30 assets, the univariate common market factor accounts, on average, for more than

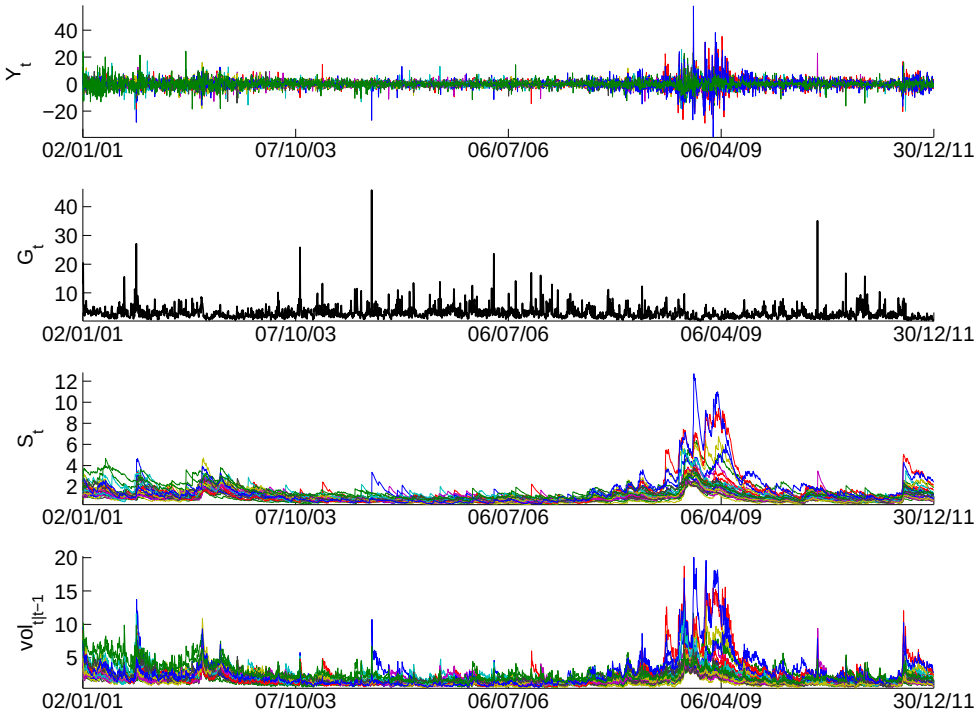

Figure 3.4: The impact of the common market factor on all of the assets. **First row:** All 30 return series. **Second row:** Values of the filtered common market factor $\widehat{G}_t$ from the ECME algorithm. **Third row:** The scale-term, $s_{k,t}$, for $k = 1, \ldots, K$, implied by the estimates of the GARCH$(1,1)$ models. **Fourth row:** The conditional volatilities of $\mathbf{Y}_t$, computed as the square root of the elements on the diagonal of matrix (4.16) and based on the parameter estimates.

40% of the conditional volatility dynamics, and the lowest 4 are (for JPM, MCD, MSFT, and WMT) around 10% to 20%. This is a consequence of separating the GARCH dynamics from the volatility shock dynamics. The former are responsible for modeling the volatility persistence. The latter are modeled by the SV dynamics of the common market factor, and capture the sharp changes in the volatility.

Figure 3.7 displays the higher-order dynamics implied by the MALap-CCC hybrid GARCH$(1,1)$-SV model and computed as in Scott et al. (2011). The first panel plots the conditional skewness. Depending on the sign of the γ_k for $k = 1, 2, \ldots, 30$, the corresponding asset exhibits either a positive or a negative skewness and its dynamics are driven by the dynamics of the G_t parameters (the correlation between $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$ and the conditional skewness is ± 0.87). The second panel displays the conditional kurtosis. It is common for all the assets because, as opposed to the conditional skewness, there is no vector which would differentiate the impact of

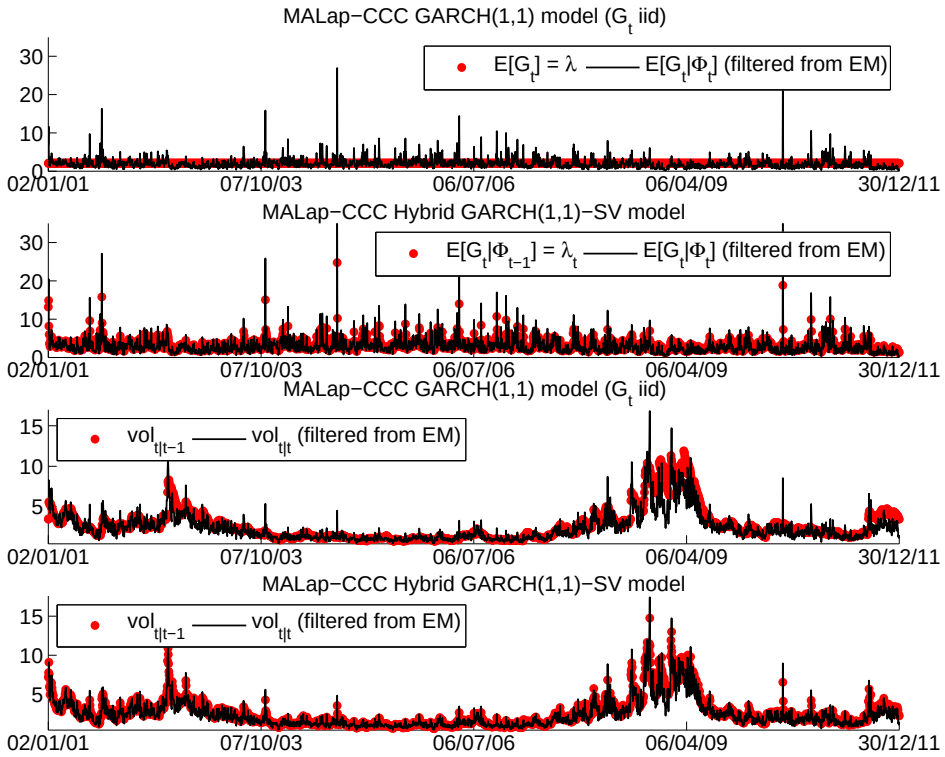

Figure 3.5: The consequences of the hybrid GARCH(1, 1)-SV extension. **First row:** The filtered $\widehat{G}_t$ values from the ECME algorithm (i.e., $\mathbb{E}[G_t \mid \mathbf{\Phi}_t]$) and the estimates obtained from the MALap-CCC GARCH(1,1) model. **Second row:** The filtered $\widehat{G}_t$ values from the ECME algorithm (i.e., $\mathbb{E}[G_t \mid \mathbf{\Phi}_t]$) and the estimates obtained from the MALap-CCC hybrid GARCH(1,1)-SV model. **Third row:** Conditional volatilities filtered from the ECME algorithm ($\text{vol}_{t|t}$) and the estimates obtained from the MALap-CCC GARCH(1,1) model ($\text{vol}_{t|t-1}$). **Fourth row:** Conditional volatilities filtered from the ECME algorithm ($\text{vol}_{t|t}$) and the estimates obtained from the MALap-CCC hybrid GARCH(1,1)-SV model ($\text{vol}_{t|t-1}$).

$\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$. From this panel and the second panel in Figure 3.5, one can note that the kurtosis and $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$ are inversely related, i.e., the lower the value of $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$, the higher the value of the kurtosis. In fact, the correlation between $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$ and the conditional kurtosis is equal to -0.81.

3.6.2 Density Forecasting Performance Comparison

Now turning to out-of-sample forecasting performance, this section compares a number of special cases of model (4.1) with the CCC model of Bollerslev (1990), the DCC model of Engle (2002), the cDCC model of Aielli (2011), and the VC model of Tse and Tsui (2002), all denoted

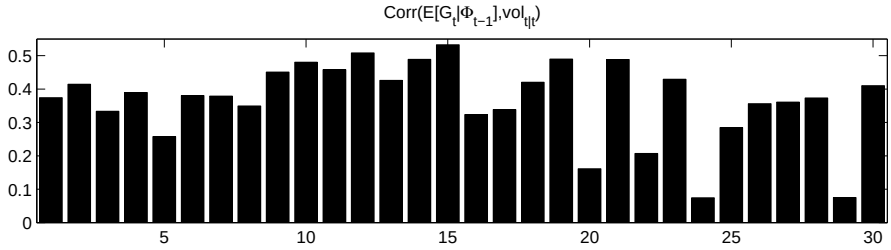

Figure 3.6: Correlation between $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}] = \lambda_t$ and conditional volatility, $\text{vol}_{t|t}$, of each of the assets filtered from the ECME algorithm (MALap-CCC hybrid GARCH(1, 1)-SV model).

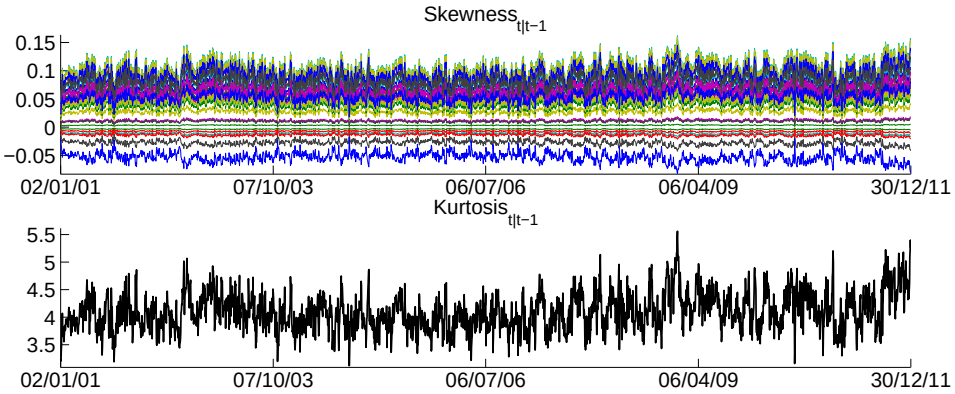

Figure 3.7: Dynamics of higher conditional moments of the returns implied by the MALap-CCC GARCH(1, 1)-SV model, computed as in Scott et al. (2011). **Upper panel:** Conditional skewness. **Bottom panel:** Conditional kurtosis.

with a prefix MN-, for multivariate normal distribution of the innovations. For each model, both GARCH(1, 1) and GJR-GARCH(1, 1) univariate dynamics are employed.

Our interest centers on the quality of one-step ahead predictions of the return vector density. For this purpose, we estimate all the models using a rolling window of 1, 000 observations, and, similar to Paolella (2013), we use the normalized sum of the realized predictive log-likelihood values, which, for given model $\mathcal{M}$, is

$$S_T(\mathcal{M}) = \frac{1}{T}\sum_{t=1}^{T} \pi_t(\mathcal{M}), \qquad \text{where} \qquad \pi_t(\mathcal{M}) = \log f^{\mathcal{M}}_{t+1|t}(\mathbf{Y}_{t+1} \mid \boldsymbol{\theta}). \tag{3.33}$$

In case of the hybrid GARCH-SV models we use, a first order approximation to π_t, and replace random parameters of G_{t+1} with the values implied by the conditional expectations $\mathbb{E}[G_{t+1} \mid \mathbf{\Phi}_t]$.

The results are given in Table 4.1. The hybrid MALap-CCC GARCH(1, 1)-SV model performs best. It is closely followed by the MNIG-CCC GARCH(1, 1)-SV model. Next in the ranking is the MAt model, followed by the MNIG and MALap models (without hybrid dynamics). The Gaussian-based models perform the worst. Interestingly, even the MALap-IID model,

$\mathcal{M}$	$S_T(\mathcal{M})$
MALap-CCC GARCH(1, 1)-SV	−46.064
NIG-CCC GARCH(1, 1)-SV	−46.064
MAt-CCC GARCH(1, 1)	−46.097
NIG-CCC GARCH(1, 1)	−46.133
MALap-CCC GARCH(1, 1)	−46.161
MALap-CCC GJR-GARCH(1, 1)-SV	−46.300
MAt-CCC GJR-GARCH(1, 1)	−46.301
NIG-CCC GJR-GARCH(1, 1)-SV	−46.336
MALap-CCC GJR-GARCH(1, 1)	−46.378
NIG-CCC GJR-GARCH(1, 1)	−46.399
MALap-IID	−47.067
Normal-cDCC GARCH(1, 1)	−47.859
Normal-DCC GARCH(1, 1)	−47.863
Normal-VC GARCH(1, 1)	−47.885
Normal-cDCC GJR-GARCH(1, 1)	−47.892
Normal-DCC GJR-GARCH(1, 1)	−47.895
Normal-VC GJR-GARCH(1, 1)	−47.914
Normal-CCC GARCH(1, 1)	−47.951
Normal-CCC GJR-GARCH(1, 1)	−47.973

Table 3.1: Performance of the one-step ahead predictions of the return vector density for different models, $\mathcal{M}$, and measured by $S_T(\mathcal{M})$, in (4.35). **First panel:** Hybrid GARCH-SV and GARCH-type models proposed in this paper. **Second panel:** MALap model under iid assumption. **Third panel:** Gaussian-based models.

without any GARCH dynamics, performs better than all Gaussian-based models, in particular, even with GARCH.

Regarding the GJR-GARCH(1, 1) dynamics, according to the results in Table 4.1, its use does not lead to better forecasting performance in any of the models. Figure 3.8 plots two tail quantiles, the means, and the medians of the estimates of the η_k, $k = 1 \ldots, 30$, from (4.9), across the moving window of 1,000 observations, for the MN-CCC GJR-GARCH model and the MALap-CCC GJR-GARCH(1, 1) model. The latter model exhibits smoother η_k estimates, and it is clearer that, in periods of high volatility such as the crisis in 2008, there was a large increase in the asymmetry effect. It thus appears that the use of GJR dynamics is enhanced, in terms of clarity and effect, when using a distribution which accounts for skewness and heavy tails.

In order to further investigate this, we check the forecasting performance of our models with the GJR-GARCH(1, 1) dynamics for the data windows when the η_k parameters are all larger than a small threshold (we use $\widehat{\eta}_k > 0.01$ for $k = 1, \ldots, 30$). It turns out that, for those

windows, and for all the distributions considered (MN, MALap, MNIG, and t), the models with GJR-GARCH$(1,1)$ significantly outperform their plain GARCH counterparts, but the improvement is much smaller than the gains obtained from relaxing the normality assumption, and from the gains associated with the hybrid GARCH-SV dynamics. In other words, the asymmetry in the volatility, captured by GJR-GARCH$(1,1)$, improves the forecasting only if it is sufficiently strongly supported by the data, and then, the improvement is small, relative to the improvements obtained by use of non-normality and the SV extension.

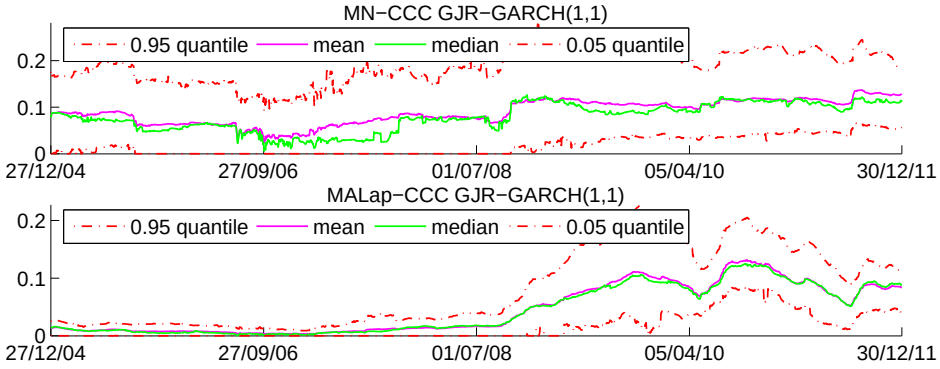

Figure 3.8: Two tail quantiles, mean, and median of η_k, $k = 1\ldots,30$ parameters from GJR-GARCH$(1,1)$ dynamics across the moving estimation window of $1,000$ observations. **Upper panel:** The MN-CCC GJR-GARCH$(1,1)$ model. **Bottom panel:** The MALap-CCC GJR-GARCH$(1,1)$ model.

The most pronounced improvement in forecasting performance is obtained when moving from the Gaussian-based models to any of the new models. The gap in forecasting performance between the new models (first panel in Table 4.1) and the Gaussian-based models (third panel in Table 4.1) is much larger than the gap between any models in a given panel.

In order to statistically test the forecasting results from Table 4.1, we use the test for unconditional predictive ability of Diebold and Mariano (2002) (see also Giacomini and White, 2006). We use a one sided test ($\mathcal{M}_1 \succ \mathcal{M}_2$) and compare each model, $\mathcal{M}_1$, in Table 4.1, with models $\mathcal{M}_2$ which resulted in a worse-than-model-$\mathcal{M}_1$ forecast. Tables of the test results are given in Paolella and Polak (2013b).

Summarizing, the first three models from Table 4.1 are very competitive and, according to the test results, there is no significant difference in forecasting performance between them. The first significant improvement (at the 5% level) occurs when moving from the MALap-CCC GARCH$(1,1)$-SV model to the MNIG-CCC GARCH$(1,1)$ model. The MALap-CCC GARCH$(1,1)$ and MNIG-CCC GARCH$(1,1)$ models perform significantly worse than the analogous hybrid models. In particular, the extension to hybrid dynamics places the MALap-CCC GARCH$(1,1)$-SV model on top. Importantly, the difference between any GARCH-type model and a corresponding hybrid GARCH$(1,1)$-SV extension is highly significant.

When moving from the Gaussian-based models to any of the new models, the t-statistic ranges from 62 to 83. In comparison, moving from a very simple MN-CCC GARCH$(1,1)$

model to the very popular and best-performing among Gaussian-based models, the MN-DCC GARCH(1, 1) model, results in a t-statistic of only 4.2. This illustrates that, even with a reasonable law of motion for the conditional volatility, the use of Gaussian innovations in such a conditional model is blatantly inferior to use of just an iid model but with a more suitable distribution. (This is not the first occurrence of such a result: It was also found using an iid model based on a two-component discrete mixture of normals, in conjunction with short estimation windows and use of shrinkage estimation; see Paolella, 2013.) In turn, using the superior distribution, in this case, the MALap, in conjunction with a GARCH structure, yields further improvement in the forecasts. In particular, comparing the MALap-CCC GARCH(1, 1)-SV to the MALap-IID model results in a t-statistic of 33.

In order to further investigate the forecasting gains from the SV extension of our model for each forecast, we use the percentage measure (defined for $\pi_t(\mathcal{M}_1)\,\pi_t(\mathcal{M}_2) > 0$)

$$D_t(\mathcal{M}_1, \mathcal{M}_2) = 100\left(|\pi_t(\mathcal{M}_1)| - |\pi_t(\mathcal{M}_2)|\right) / |\pi_t(\mathcal{M}_2)|\,. \tag{3.34}$$

In Figure 3.9, we plot D_t for the MALap-CCC hybrid GARCH(1, 1)-SV and the MALap-CCC GARCH(1, 1). We find that (i) on average, the SV extension results in only a minor improvement in forecasting performance even when we consider only periods of large average absolute returns; (ii) but when compared across time, forecasts during the period of the 2008 crisis display a systematic improvement from the SV extension.

3.6.3 Mean Forecasts

Lastly, we consider the forecast of the mean; this being, for example, of utmost importance in a portfolio selection context; see, e.g., Chopra and Ziemba (1993). Figure 3.10 compares the forecasts of the conditional means based on (i) the sample mean and median from a rolling window of 1,000 observations; (ii) the model-based mean from the MALap-CCC GARCH(1, 1) model; and (iii) that from the MALap-CCC GARCH(1, 1)-SV extension. Around the 2008 crisis, the sample mean estimates are strongly influenced by negative returns, and, in general, with heavy-tailed data, the sample mean is not the optimal estimator. The sample median is more robust and, as the thickness of the tail increases, it becomes a more efficient estimator. Indeed, the MALap-CCC GARCH(1, 1) mean forecasts are more similar to the median estimates. This exercise helps confirm that the model-based forecasts of the mean are accurate.

A potential drawback of the hybrid GARCH(1, 1)-SV model is that the dynamics in (4.4) have an impact on mean dynamics. The forecasts based on the MALap-CCC GARCH(1, 1)-SV model, given in the last panel of Figure 3.10, are more varying, because the conditional mean is a function of the $G_t \mid \mathbf{\Phi}_{t-1}$ parameters as in (4.15). One could consider more general SV dynamics incorporating the moving average component into (4.4). This would smooth the forecasts in the last panel of Figure 3.10 and result in further improvement of the forecasting performance of the hybrid GARCH(1, 1)-SV models. To investigate this, we modified the forecast conditional density by scaling the estimate of $\boldsymbol{\gamma}$ with the factor $\{\widehat{c}/(1-\widehat{\rho})\}/\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$, where the hatted values come from estimation. This has the effect of removing the impact of the spikes in

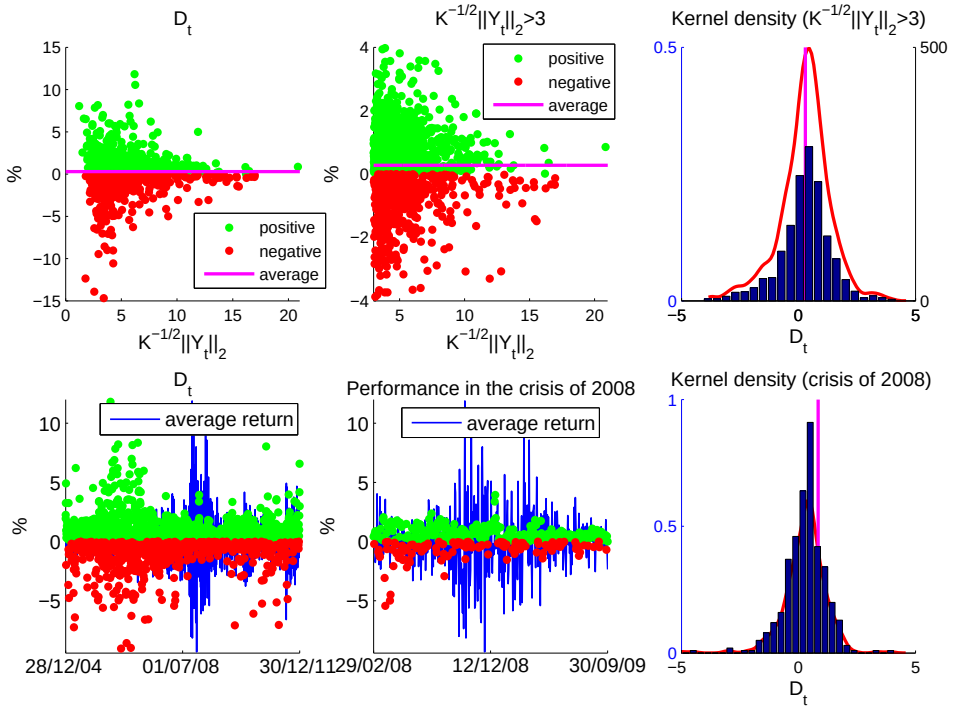

Figure 3.9: The forecasting gains from the hybrid GARCH$(1,1)$-SV extension. **First row, column-wise:** Percentage gains D_t from (3.34) as a function of average absolute return. Using MALap-CCC GJR-GARCH$(1,1)$ as $\mathcal{M}_1$ and MALap-CCC GARCH$(1,1)$ as $\mathcal{M}_2$. Same, but for large average absolute returns. Histogram of percentage gains for large average absolute returns. **Second row, column-wise:** Percentage gains D_t from (3.34) as a function of time. Same, but for crisis of 2008. Histogram of percentage gains during the 2008 crisis.

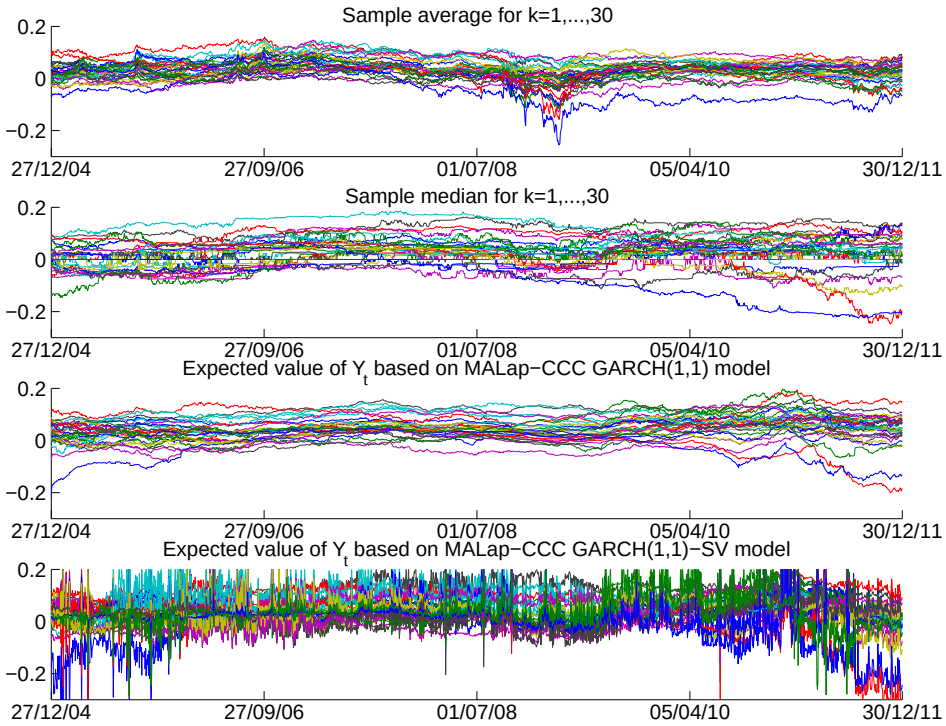

Figure 3.10: Conditional mean forecasts from a rolling window of $1,000$ observations. **First row:** Sample mean. **Second row:** Sample median. **Third row:** The MALap-CCC GARCH$(1,1)$ model. **Fourth row:** The MALap-CCC GARCH$(1,1)$-SV model.

$\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}]$ in the mean equation (4.15). This results in improved forecasting performance, but it was not statistically significant at the 5% level.

3.7 Conclusions

We introduce a new class of models which combines GARCH-type dynamics with an SV structure (hybrid GARCH-SV class). The former captures the asset-specific volatility clustering effects and the latter is responsible for common market shocks. The proposed model also allows for a new type of dynamic in the dependency structure leading to additional dynamics in the higher-order moments. Maximum likelihood estimation is numerically reliable and fast, and can be used with a large number of assets. It yields consistent and asymptotically normal estimates of the parameters. In- and out-of-sample exercises provide justification for use of the model with real data. The model delivers a non-Gaussian predictive distribution with a tractable sum of margins, and hence can be straightforwardly applied to portfolio optimization.

The model also lends itself to multivariate option pricing by combining the equivalent martingale measure technique in the presence of a GARCH structure with an iterative estimation of the model dynamics. Future work will pursue the theoretical properties, convergence analysis, and empirical performance of the proposed option pricing algorithm.

In future research, one could entertain the use of the multivariate noncentral t distribution, as used in Jondeau (2012). This also arises as a mixture similar to (4.1) but such that, in (4.1a), the square root of G_t is used. As such, the noncentral t is not a special case of the MGHyp, but is still such that its sums of marginals is noncentral t, and has the appealing property of asymmetric tail dependence.

The dependency matrix $\mathbf{\Gamma}$ could be augmented by information in high-frequency data, possibly along the lines of Noureldin et al. (2012). Finally, the CCC structure could be modified in the same vein as in the DCC model of Engle (2002), the VC model of Tse and Tsui (2002), and the switching CCC model of Pelletier (2006). These models, particularly the latter, will entail substantially more estimation time, but might yield improved density forecasts. These ideas are currently being pursued.

3.8 Appendices

3.8.1 Link to the Taylor (1982) SV model

The univariate model proposed in Taylor (1982) is given by

$$Y_t = \exp(Q_t/2)\, Z_t, \quad \text{with} \quad Q_t = c + \rho Q_{t-1} + \sigma \eta_{t-1}, \tag{3.35}$$

for $Z_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$; $\eta_t \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$; and Z_t and η_t are independent. Define Q_t by $G_t = \exp(Q_t)$. Then (4.1) can be rewritten as

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\gamma} \exp(Q_t) + \mathbf{H}_t^{1/2} \exp(Q_t/2)\, \mathbf{Z}_t. \tag{3.36}$$

Switching from $(G_t \mid \mathbf{\Phi}_{t-1})$ dynamics in (4.4) to $(Q_t \mid \mathbf{\Phi}_{t-1})$ dynamics and dropping the conditional expectations, we get

$$(Q_t^r \mid \mathbf{\Phi}_{t-1}) = c_r + \rho_r \left(Q_{t-1}^r \mid \mathbf{\Phi}_{t-2}\right) + \tilde{\zeta}_{r,t}, \tag{3.37}$$

where $\tilde{\zeta}_{r,t} = (Q_t^r \mid \mathbf{\Phi}_t) - (Q_t^r \mid \mathbf{\Phi}_{t-1})$. Setting $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\gamma} = \mathbf{0}$, $r = 1$ and $\mathbf{H}_t = 1$ in (3.36) (for $K = 1$) results in (3.36) reducing to (3.35), with η_{t-1} replaced by $\tilde{\zeta}_{1,t}$ and $(Q_t \mid \mathbf{\Phi}_{t-1})$ instead of Q_t.

Our model differs from the SV model in three additional aspects. Firstly, we replace the past shock η_{t-1} by the current shock η_t in (3.35). In the SV literature, one has to use a lag shock (together with $\mathrm{Corr}(Z_t, \eta_t) < 0$) to obtain the key feature of SV models, the asymmetric return-volatility relation, often called a statistical leverage effect. In our setup, we can incorporate the asymmetry or the leverage effect through the scale-term dynamics.

Note that, if we were to use $\zeta_{r,t-1}$ instead of $\zeta_{r,t}$ in our model, then there would be a one-period shift between the filtered $\widehat{G}_t$ values, and the moments of G_t conditional on $\mathbf{\Phi}_{t-1}$. Use of

$\zeta_{r,t}$ avoids this and allows for shocks to the volatility to have an immediate impact on returns, a feature which is absent in discrete time SV models.

The second difference is: we work with $G_t \mid \mathbf{\Phi}_{t-1}$ instead of $\log G_t \mid \mathbf{\Phi}_{t-1}$ because the former has tractable moment expressions, and we can still guarantee positive values of $\mathbb{E}\left[G_t \mid \mathbf{\Phi}_{t-1}\right]$ by imposing the constraints $c_r > 0$ and $\rho_r \geq 0$.

Thirdly, the dynamics in our model are written in terms of $\mathbb{E}\left[G_t \mid \mathbf{\Phi}_{t-1}\right]$ because G_t is a latent process. Here, we are able to filter them out through our ECME algorithm without an additional computational burden. More importantly, keeping the dynamics of G_t only in terms of conditional expectations allows us to maintain, without any extra conditions, the monotonic increase of the incomplete-data likelihood which is a key property of the ECME algorithm.

3.8.2 Derivation of the Q^{+G}-Dynamics for Option Pricing

Following Christoffersen et al. (2010) and a multivariate extension given in Rombouts and Stentoft (2011), we impose the exponential affine form on the Radon-Nikodym derivative, with respect to $\mathcal{F}_t^{+G}$. Hence, by the law of iterative expectation, we get

Lemma 3.8.1. *For any K-dimensional sequence $\mathbf{v}_s$,*

$$\frac{dQ^{+G}}{dP} \mid \mathcal{F}_t^{+G} = \exp\left(-\sum_{s=1}^{t}\left(\mathbf{v}_s'\boldsymbol{\varepsilon}_s + \frac{1}{2}\mathbf{v}_s'\mathbf{H}_s\mathbf{v}_s G_s\right)\right) \tag{3.38}$$

is a Radon-Nikodym derivative with respect to $\mathcal{F}_t^{+G}$.

Proof. We need to show that $\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_t^{+G} > 0$ and $\mathbb{E}_0^P\left[\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_t^{+G}\right] = 1$. Non-negativity is an immediate consequence of the exponential form. For the second condition we use the law of iterated expectations, with respect to $\mathcal{F}_t^{+G}$, to obtain

$$\begin{aligned}
\mathbb{E}_0^P\left[\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_t^{+G}\right] &= \mathbb{E}_0^P\left[\exp\left(-\sum_{s=1}^{t}\left(\mathbf{v}_s'\boldsymbol{\varepsilon}_s + \frac{1}{2}\mathbf{v}_s'\mathbf{H}_s\mathbf{v}_s G_s\right)\right)\right] \\
&= \mathbb{E}_0^P\left[\mathbb{E}_1^P, \ldots, \mathbb{E}_{t-1}^P \exp\left(\sum_{s=1}^{t}\left(-\mathbf{v}_s'\boldsymbol{\varepsilon}_s\right) - \sum_{s=1}^{t}\frac{1}{2}\mathbf{v}_s'\mathbf{H}_s\mathbf{v}_s G_s\right)\right] \\
&= \mathbb{E}_0^P\left[\mathbb{E}_1^P, \ldots, \mathbb{E}_{t-2}^P \exp\left(\sum_{s=1}^{t-1}\left(-\mathbf{v}_s'\boldsymbol{\varepsilon}_s\right) - \sum_{s=1}^{t}\frac{1}{2}\mathbf{v}_s'\mathbf{H}_s\mathbf{v}_s G_s\right)\mathbb{E}_{t-1}^P \exp\left(-\mathbf{v}_t'\boldsymbol{\varepsilon}_t\right)\right] \\
&= \mathbb{E}_0^P\left[\mathbb{E}_1^P, \ldots, \mathbb{E}_{t-2}^P \exp\left(\sum_{s=1}^{t-1}\left(-\mathbf{v}_s'\boldsymbol{\varepsilon}_s\right) - \sum_{s=1}^{t}\frac{1}{2}\mathbf{v}_s'\mathbf{H}_s\mathbf{v}_s G_s\right)\mathbb{E}_{t-1}^P \exp\left(\frac{1}{2}\mathbf{v}_t'\mathbf{H}_t\mathbf{v}_t G_t\right)\right] \\
&= \mathbb{E}_0^P\left[\mathbb{E}_1^P, \ldots, \mathbb{E}_{t-2}^P \exp\left(\sum_{s=1}^{t-1}\left(-\mathbf{v}_s'\boldsymbol{\varepsilon}_s\right) - \sum_{s=1}^{t-1}\frac{1}{2}\mathbf{v}_s'\mathbf{H}_s\mathbf{v}_s G_s\right)\right],
\end{aligned}$$

where the second last equality follows from the normality of $\boldsymbol{\varepsilon}_t$ conditional on $\mathcal{F}_{t-1}^{+G}$. Iterating on this yields the required result. □

Having a valid candidate for the change of measure, we proceed to find conditions for the sequence $\mathbf{v}_s$ under which the proposed Radon-Nikodym derivative defines an EMM under $\mathcal{F}_t^{+G}$. Denote by $\mathbf{r} = (r, r, \ldots, r)'$ a K vector of risk free interest rates, then

Proposition 3.8.1. *The probability measure Q^{+G} defined by the Radon-Nikodym derivative in* (3.38) *is an EMM under $\mathcal{F}_t^{+G}$ if and only if*

$$\mathbf{v}_t = \mathbf{H}_t^{-1}\left(\left(\boldsymbol{\mu} - \mathbf{r}\right) G_t^{-1} + \boldsymbol{\gamma} + \frac{1}{2}\mathrm{diag}\left(\mathbf{S}_t^2\right)\right).$$

Proof. We need to show that, for all $k = 1, 2, \ldots, K$, $\mathbb{E}^{Q^{+G}}\left[\frac{X_{t,k}}{X_{t-1,k}} \mid \mathcal{F}_{t-1}^{+G}\right] = \exp\left(r\right)$, where r is the risk free interest rate, and $X_{t,k}$ is the price of stock k at time t. We have

$$\begin{aligned}
&\mathbb{E}^{Q^{+G}}\left[\frac{X_{t,k}}{X_{t-1,k}}\exp\left(-r\right) \mid \mathcal{F}_{t-1}^{+G}\right] = \\
&= \mathbb{E}^{P}\left[\left(\frac{\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_t^{+G}}{\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_t}\right)\left(\frac{\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_t}{\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_{t-1}^{+G}}\right)\frac{X_{t,k}}{X_{t-1,k}}\exp\left(-r\right) \mid \mathcal{F}_{t-1}^{+G}\right] \\
&= \mathbb{E}^{P}\left[\left(\frac{\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_t^{+G}}{\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_{t-1}^{+G}}\right)\frac{X_{t,k}}{X_{t-1,k}}\exp\left(-r\right) \mid \mathcal{F}_{t-1}^{+G}\right] \\
&= \mathbb{E}^{P}\left[\exp\left(-\mathbf{v}_t'\boldsymbol{\varepsilon}_t - \frac{1}{2}\mathbf{v}_t'\mathbf{H}_t\mathbf{v}_t G_t\right)\exp\left(\mu_k + \gamma_k G_t + \varepsilon_{t,k}\right)\exp\left(-r\right) \mid \mathcal{F}_{t-1}^{+G}\right] \\
&= \mathbb{E}^{P}\left[\exp\left(-\frac{1}{2}\mathbf{v}_t'\mathbf{H}_t\mathbf{v}_t G_t + \mu_k + \gamma_k G_t - r - \mathbf{v}_t'\boldsymbol{\varepsilon}_t + \mathbf{e}_k'\boldsymbol{\varepsilon}_{\boldsymbol{t}}\right) \mid \mathcal{F}_{t-1}^{+G}\right] \\
&= \exp\left(-\frac{1}{2}\mathbf{v}_t'\mathbf{H}_t\mathbf{v}_t G_t + \mu_k + \gamma_k G_t - r + \frac{1}{2}\left(\mathbf{v}_t - \mathbf{e}_k\right)'\mathbf{H}_t\left(\mathbf{v}_t - \mathbf{e}_k\right)G_t\right). \\
&= \exp\left(-\mathbf{e}_k'\mathbf{H}_t\mathbf{v}_t G_t + \frac{1}{2}\mathbf{e}_k'\mathbf{H}_t\mathbf{e}_k G_t + \mu_k + \gamma_k G_t - r\right),
\end{aligned}$$

where $\mathbf{e}_k = (0, \ldots, 0, 1, 0, \ldots, 0)'$ a vector of zeros with one at position k. Thus, if we ensure that

$$-\mathbf{e}_k'\mathbf{H}_t\mathbf{v}_t G_t + \frac{1}{2}\mathbf{e}_k'\mathbf{H}_t\mathbf{e}_k G_t + \mu_k + \gamma_k G_t - r = 0$$

for all $k = 1, 2, \ldots, K$, by choosing the vector series $\mathbf{v}_t$, then the probability measure Q^{+G} is an EMM since it makes discounted asset prices martingales. Solving it for $\mathbf{v}_t$, in vector form we obtain

$$\mathbf{v}_t = \mathbf{H}_t^{-1}\left(\left(\boldsymbol{\mu} + \boldsymbol{\gamma}G_t - \mathbf{r}\right) G_t^{-1} + \frac{1}{2}\mathrm{diag}\left(\mathbf{S}_t^2\right)\right).$$

□

Thus, if we would know the realizations of G_s for $s = 1, 2, \ldots, t$, then it is possible to solve explicitly for the respective $\mathbf{v}_s$, and Proposition 3.8.1 guarantees that the corresponding measure is an EMM under $\mathcal{F}_t^{+G}$.

Use of the $\mathcal{F}_t^{+G}$ filtration implies that, although we are working within an incomplete market framework, there is only one source of randomness. Therefore, constraining the Radon-Nikodym

derivative to be of the exponential affine form, as in (3.38), allows us to derive a unique measure under which the discounted asset prices are Q^{+G}-martingales. Moreover, because $\mathbf{Y}_t \mid G_t$ is Gaussian, we can characterize the change of measure and the Q^{+G}-dynamics corresponding to model (4.1) explicitly. Denote by $\Psi_t(\mathbf{u})$ the logarithm of the conditional moment generating function of $\boldsymbol{\varepsilon}_t$ given $\mathcal{F}_{t-1}^{+G}$, i.e.,

$$\Psi_t(\mathbf{u}) = \frac{1}{2}\mathbf{u}'\mathbf{H}_t\mathbf{u}G_t. \tag{3.39}$$

In order to obtain the Q^{+G}-dynamics of our model, we derive the analogue of Ψ_t under Q.

Corollary 3.8.1. *The logarithm of the conditional (on $\mathcal{F}_{t-1}^{+G}$) moment generating function of $\boldsymbol{\varepsilon}_t$ under Q^{+G} is given by*

$$\Psi_t^{Q^{+G}}(\mathbf{u}) = \log \mathbb{E}^{Q^{+G}}\left[\exp\left(-\mathbf{u}'\boldsymbol{\varepsilon}_t\right) \mid \mathcal{F}_{t-1}^{+G}\right] = \Psi_t(\mathbf{v}_t + \mathbf{u}) - \Psi_t(\mathbf{v}_t). \tag{3.40}$$

Proof. By change of measure and rearranging we get

$$\begin{aligned}
\mathbb{E}^{Q^{+G}}\left[\exp\left(-\mathbf{u}'\boldsymbol{\varepsilon}_t\right) \mid \mathcal{F}_{t-1}^{+G}\right] &= \mathbb{E}^{P}\left[\left(\frac{\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_t^{+G}}{\frac{\mathrm{d}Q^{+G}}{\mathrm{d}P} \mid \mathcal{F}_{t-1}^{+G}}\right)\exp\left(-\mathbf{u}'\boldsymbol{\varepsilon}_t\right) \mid \mathcal{F}_{t-1}^{+G}\right] \\
&= \mathbb{E}^{P}\left[\exp\left(-\mathbf{v}_t'\boldsymbol{\varepsilon}_t - \Psi_t(\mathbf{v}_t)\right)\exp\left(-\mathbf{u}'\boldsymbol{\varepsilon}_t\right) \mid \mathcal{F}_{t-1}^{+G}\right] \\
&= \mathbb{E}^{P}\left[\exp\left(-\left(\mathbf{v}_t + \mathbf{u}\right)'\boldsymbol{\varepsilon}_t - \Psi_t(\mathbf{v}_t)\right) \mid \mathcal{F}_{t-1}^{+G}\right] \\
&= \exp\left(\Psi_t(\mathbf{v}_t + \mathbf{u}) - \Psi_t(\mathbf{v}_t)\right).
\end{aligned}$$

Taking log of both sides completes the proof. □

Substituting (3.39) into (3.40), we get

$$\Psi_t^{Q^{+G}}(\mathbf{u}) = \mathbf{u}'\mathbf{H}_t\mathbf{v}_tG_t + \frac{1}{2}\mathbf{u}'\mathbf{H}_t\mathbf{u}G_t.$$

Now, using the expression for $\mathbf{v}_t$ given in Proposition 3.8.1,

$$\Psi_t^{Q^{+G}}(\mathbf{u}) = \mathbf{u}'(\boldsymbol{\mu} - r) + \mathbf{u}'\boldsymbol{\gamma}G_t + \frac{1}{2}\mathbf{u}'\mathrm{diag}\left(\mathbf{S}_t^2\right)G_t + \frac{1}{2}\mathbf{u}'\mathbf{H}_t\mathbf{u}G_t. \tag{3.41}$$

So, the Q^{+G}-dynamics of the returns remain Gaussian with a shift in the mean, as in (3.30).

3.8.3 Iterative Change of Measure (ICM) Algorithm

If the returns follow the dynamics from the COMFORT model (4.1), then one can construct a sequence of measures $Q^{[\ell]}$ by the following algorithm. It uses the historical measure P estimates as the starting values, and iterates the following steps, for $\ell = 1, 2, \ldots$

E-step: Calculate $\mathbb{E}^{Q^{[\ell]}}\left[\log L_{\mathbf{Y},\mathbf{G}} \mid \mathbf{Y}; \widehat{\boldsymbol{\theta}}_P^{Q^{[\ell]}}, \widehat{\boldsymbol{\theta}}_D^{Q^{[\ell]}}, \widehat{\boldsymbol{\theta}}_C^{Q^{[\ell]}}\right]$. Computationally, this step is the same as under historical measure P; the difference is only in the distribution parameters.

CM1-step: Update $\boldsymbol{\theta}_P$, $\boldsymbol{\theta}_C$, i.e., Gaussian parameters under measure P, with the unobserved realizations of G_t's replaced by their filtered values, under $Q^{[\ell]}$, from the E-step above.

(P) Update $\boldsymbol{\theta}_P$ by computing

$$\arg\max_{\boldsymbol{\theta}_P} \log L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}^{Q^{[\ell]}}}(\boldsymbol{\theta}_P), \tag{3.42}$$

where $L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}^{Q^{[\ell]}}}(\boldsymbol{\theta}_P)$ is a Gaussian likelihood with zero correlation, so we can estimate the parameters of each asset, $(\mu_k, \gamma_k, \omega_k, \alpha_k, \beta_k)$, separately by maximizing the corresponding likelihood function.

(C) Update $\boldsymbol{\theta}_C$ by computing the usual empirical correlation estimator (the MLE under normality) of the de-volatilized residuals $\left(\widehat{G}_t^{Q^{[\ell]}}\right)^{-1/2} \mathbf{S}_t^{-1}\widehat{\boldsymbol{\varepsilon}}_t$, where $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\gamma}}\widehat{G}_t^{Q^{[\ell]}}$, $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\gamma}}$ are obtained in part (P) directly above, and $\widehat{G}_t^{Q^{[\ell]}}$ is obtained in the E-step.

EMM-step Given the CM1-step estimates of $\boldsymbol{\theta}_P$, $\boldsymbol{\theta}_C$, and the filtered, under measure $Q^{[\ell]}$ values of G_t, change the measure as detailed in Appendix 3.8.2, and obtain the estimates of $\boldsymbol{\theta}_P^{Q^{[\ell+1]}}$, $\boldsymbol{\theta}_C^{Q^{[\ell+1]}}$.

CM2-step: Given the estimates of $\boldsymbol{\theta}_P^{Q^{[\ell+1]}}$, $\boldsymbol{\theta}_C^{Q^{[\ell+1]}}$, obtain new estimates of $\boldsymbol{\theta}_D$, under $Q^{[\ell+1]}$, by maximizing the incomplete data log-likelihood function, i.e., compute

$$\widehat{\boldsymbol{\theta}}_D^{Q^{[\ell+1]}} = \arg\max_{\boldsymbol{\theta}_D} \log L_{\mathbf{Y}}\left(\boldsymbol{\theta}_D \mid \widehat{\boldsymbol{\theta}}_P^{Q^{[\ell+1]}}, \widehat{\boldsymbol{\theta}}_C^{Q^{[\ell+1]}}\right), \tag{3.43}$$

where

$$\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1} \overset{Q^{[\ell+1]}}{\sim} \mathrm{MGHyp}\left(\widehat{\boldsymbol{\mu}}^{Q^{[\ell+1]}}, \widehat{\boldsymbol{\gamma}}^{Q^{[\ell+1]}}, \mathbf{H}_t^{Q^{[\ell+1]}}, \boldsymbol{\theta}_D\right), \tag{3.44}$$

and $\mathbf{H}_t^{Q^{[\ell+1]}} = \mathbf{S}_t^{Q^{[\ell+1]}} \boldsymbol{\Gamma}^{Q^{[\ell+1]}} \mathbf{S}_t^{Q^{[\ell+1]}}$ is the dispersion matrix under $Q^{[\ell+1]}$, with the scale term dynamics, in a diagonal matrix $\mathbf{S}_t^{Q^{[\ell+1]}}$, with the $\widehat{\boldsymbol{\theta}}_P^{Q^{[\ell+1]}}$ parameters from the EMM step above.

Iterate the above steps until

$$\left|\log L_{\mathbf{Y}}\left(\widehat{\boldsymbol{\theta}}_P^{Q^{[\ell+1]}}, \widehat{\boldsymbol{\theta}}_C^{Q^{[\ell+1]}}, \widehat{\boldsymbol{\theta}}_D^{Q^{[\ell+1]}}\right) - \log L_{\mathbf{Y}}\left(\widehat{\boldsymbol{\theta}}_P^{Q^{[\ell]}}, \widehat{\boldsymbol{\theta}}_C^{Q^{[\ell]}}, \widehat{\boldsymbol{\theta}}_D^{Q^{[\ell]}}\right)\right| \leq \varepsilon,$$

for some fixed $\varepsilon > 0$.

3.8.4 Evaluation of the Bessel Function

Let $G \sim \mathrm{GIG}(\lambda, \chi, \psi)$, for $\chi > 0$ and $\psi > 0$. Then it may be shown (see, e.g., Paolella, 2007, Ch. 9) that

$$\mathbb{E}[G^\alpha] = \left(\frac{\chi}{\psi}\right)^{\alpha/2} \frac{K_{\lambda+\alpha}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}, \quad \alpha \in \mathbb{R}, \tag{3.45}$$

which involves a ratio of Bessel functions $K_v(z)$ as given in (4.3). For large z ($z > 700$) and $v < 55$, the Matlab function *besselk.m* returns 0 and hence the ratio is undefined.

It is possible to compute the limit of the Bessel function ratio for some cases. Let $\mathbf{Y} \mid (G = g) \sim \mathrm{N}(\boldsymbol{\mu} + \boldsymbol{\gamma} g, g\boldsymbol{\Sigma})$. We are interested in the expectations of $G^{\pm 1} \mid (\mathbf{Y} = \mathbf{y})$ which, for $\psi \neq 0$ or $\gamma \neq 0$, are always positive and have their limits given by

$$\lim_{m \to \infty} \left(\frac{m + \chi}{\psi + \boldsymbol{\gamma} \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}'} \right)^{\pm 1/2} \frac{K_{\lambda - K/2 \pm 1} \left(\sqrt{(m + \chi)\left(\psi + \boldsymbol{\gamma} \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}'\right)} \right)}{K_{\lambda - K/2} \left(\sqrt{(m + \chi)\left(\psi + \boldsymbol{\gamma} \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}'\right)} \right)} = 0,$$

where $m = (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$.

In our model, these ratios are responsible for proper weights, in the E-step and CM1-step of the ECME algorithm. We thus require a highly accurate approximation of Bessel function ratios for large v or z. This can be done by using the Watson (1922, p. 202) asymptotic expansion of $K_v(z)$ given by

$$K_v(z) = \sqrt{\frac{\pi}{2z}} \exp(-z) E(v, z), \tag{3.46}$$

where

$$E(v, z) = 1 + \sum_{k=1}^{\infty} \frac{\prod_{l=1}^{k} \left(4v^2 - (2l-1)^2\right)}{k! (8z)^k}. \tag{3.47}$$

Inspection of (3.46) reveals that, for a ratio of Bessel functions, a numerically problematic $\exp(z)$ cancels out. In order to use (3.46), we have to truncate the series at some finite K which causes, for z small ($z < 10$), some loss of accuracy (compared to the Matab function *besselk.m*), but as z increases, the accuracy grows very rapidly because of z to the k power in terms of the series (3.47).

4

Sharpening Sharpe

Sharpening Sharpe: Using Conditional Measures with non-Gaussian DCC Models for Improved Portfolio Performance

Paweł Polak[a,b]

[a] *Department of Banking and Finance, University of Zurich, Switzerland*
[b] *Swiss Finance Institute*

Abstract

It is now well-known that the dependency between assets is not constant over time. This has resulted in several models which can incorporate different dependency dynamics. Among them, the dynamic correlation models VC and DCC command the most attention in the literature because of their applicability to a large number of assets, ease of estimation, and improved model fit compared to a constant conditional correlation model. For multivariate density forecasting, risk forecasting, and portfolio allocation, a severe drawback of these models is the underlying assumption of Gaussianity for the innovation sequence. This paper proposes a model in which the innovations are multivariate generalized hyperbolic, or special cases thereof, such as variance-gamma, asymmetric Student's t, and normal inverse Gaussian. This paper is unique in that estimation of model parameters is conducted by joint maximum likelihood of all the model parameters, using an EM algorithm, so that it is still feasible for hundreds of assets, and more accurate than ad-hoc two-step methods. Besides showing improved density forecasts and risk prediction, performance of expected-shortfall-optimized portfolios is investigated. The new model class outperforms the Gaussian based models in all regards.

Keywords:Expected Shortfall, Multivariate Generalized Hyperbolic Distribution; Non-Ellipticity; Portfolio Optimization.

JEL Classification: C51; C53; G11; G17.

4.1 Introduction

The Constant Conditional Correlations (CCC) model of Bollerslev (1990), the Varying Correlation (VC) model of Tse and Tsui (2002), and the Dynamic Conditional Correlation (DCC) model of Engle (2002), together with their various extensions, have become among the most popular tools for modeling and predicting multivariate asset volatility dynamics; see, in particular, Engle (2009) and the references therein. Part of the reason for their success is their

applicability to large dimensional data—a feature not shared by several other classes of multivariate GARCH models; see, for example, the discussion in Paolella and Polak (2013a) and the references therein.

To some extent, the DCC model has become a benchmark and industry standard, owing to its applicability to a large number of assets, its ease of programming and computation, availability in econometric software, and that it allows for at least a simple law of motion for the correlations in the otherwise constant setting of the CCC model. Nevertheless, the model is not without criticism. Somewhat trivially, the law of motion for the correlations is very simple, using only two additional parameters, and globally applies to all the assets; however this is arguably one of its strong points, and also is the reason why estimation is straightforward, even with a large number of assets. There are, however, other critiques which are fundamentally more serious, such as possible lack of consistency. A discussion of some of the flaws of DCC is given in Caporin and McAleer (2013). While acknowledging these criticisms, we take a somewhat more pragmatic approach: As stated in Caporin and McAleer (2013), "DCC may be a useful filter or a diagnostic check, but it is not a model". When viewed as a "filter", as opposed to being a model which is undoubtedly not the true DGP, DCC can have value in forecasting, because it still can pick up some of the signal in the true underlying process. Our out-of-sample density forecasting exercises confirm that, when compared to a CCC structure, DCC has merit in this regard. In particular, this is the case when using the Gaussian distributional assumption, as well as what we propose here, the non-Gaussian case. For the latter, the out-of-sample performance is demonstrably better, and the gains in forecasting accuracy attributed to a change in the distributional assumption strongly outweigh those from moving from the CCC to the DCC structure, in agreement with other studies, e.g., Paolella (2013).

In this paper, we propose models which, among other things, generalize the VC and DCC models to support the multivariate generalized hyperbolic distribution, of which Gaussianity (and numerous popular heavy-tailed distributions) are special or limiting cases. Thus, in addition to capturing the primary stylized facts of volatility clustering and changing correlations, the model also is able to address the non-Gaussianity of the innovation process, both in terms of excess kurtosis, and asymmetry. The model incorporates a further structure which gives rise to conditional skewness dynamics, a stylized fact which is now well-documented; see Jondeau and Rockinger (2005), Jondeau (2012), Haas et al. (2004) and Paolella and Polak (2013b), and the references therein.

The approach originates from the so-called COMFORT model of Paolella and Polak (2013b) (hereafter PP13), in which the set of asset returns follows a conditional multivariate generalized hyperbolic (MGHyp) distribution with a constant dependency matrix. While in many respects, this is an attractive parametrization, we show herein that, similar to the original CCC model, the assumption of a constant dependency matrix is often too restrictive. This paper investigates the extension of that model to support several DCC-like structures. In particular, we investigate use of the DCC, the VC, and the DCC bias-corrected, cDCC, model of Aielli (2011). Anticipating our empirical results below, we find that out-of-sample forecasting ability is indeed enhanced by their incorporation. It is important to emphasize that the various model extensions

considered herein all inherit the following important benefits of the CCC model in PP13: (i) fast, joint-parameter maximum likelihood estimation (MLE) via an expectation-maximization (EM) algorithm; (ii) the possibility of using parallel computing for further computational speed; (iii) the predictive distribution is a conditional MGHyp (or one of various special cases of it), so that, recalling that the sums of the margins of a MGHyp is itself a GHyp, portfolio construction is straightforward; see below for details.

The remainder of this paper is as follows. Section 4.2 states the proposed model, while Section 4.3 details its estimation. Section 4.4 presents a new result which is of great numeric value for quickly computing the expected shortfall of a generalized hyperbolic random variable. Section 4.5 presents a detailed empirical study of the comparative performance in terms of in-sample fit and out-of-sample density forecasting, risk prediction, and portfolio performance. An Appendix collects some technical results.

4.2 Model

We consider a set of K financial assets with the corresponding return vector at time t, denoted by $\mathbf{Y}_t = (Y_{t,1}, Y_{t,2}, \ldots, Y_{t,K})'$, $t = 1, 2, \ldots, T$. The equally spaced (ignoring the weekend effect for daily data) realization of the return vector is denoted by $\mathbf{Y} = [\mathbf{Y}_1 \mid \mathbf{Y}_2 \mid \cdots \mid \mathbf{Y}_T]$. We focus on modeling the dynamics of the conditional multivariate density of $\mathbf{Y}_t$ given the history of returns, $\mathbf{\Phi}_{t-1} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_{t-1}\}$. In particular, we assume that Y_t follows a conditional MGHyp distribution with the representation from PP13 given by

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\gamma} G_t + \boldsymbol{\varepsilon}_t, \quad \text{with} \tag{4.1a}$$

$$\boldsymbol{\varepsilon}_t = \mathbf{H}_t^{1/2} \sqrt{G_t} \mathbf{Z}_t, \tag{4.1b}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$ are column vectors in $\mathbb{R}^K$; $\mathbf{H}_t$ is a positive definite, symmetric, dispersion matrix of order K; $\mathbf{Z}_t \overset{\text{iid}}{\sim} \mathrm{N}(\mathbf{0}, \mathbf{I}_K)$ is a sequence of independent and identically distributed (iid) normal random variables and $(G_t \mid \mathbf{\Phi}_{t-1}) \sim \mathrm{GIG}(\lambda_t, \chi_t, \psi_t)$ are mixing random variables, $t = 1, 2, \ldots, T$, independent of $\mathbf{Z}_t$, with typical GIG (generalized inverse Gaussian) density given by

$$f_G(x; \lambda, \chi, \psi) = \frac{\chi^{-\lambda} \left(\sqrt{\chi\psi}\right)^{\lambda}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} x^{\lambda-1} \exp\left(-\frac{1}{2}\left(\chi x^{-1} + \psi x\right)\right), \quad x > 0; \tag{4.2}$$

$K_\lambda(x)$ is the modified Bessel function of the third kind (and not to be confused with K, the number of assets), given by

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty t^{\lambda-1} \exp\left(-\frac{x}{2}\left(t + t^{-1}\right)\right) \mathrm{d}t, \quad x > 0; \tag{4.3}$$

and $\chi_t > 0$, $\psi_t \geq 0$ if $\lambda_t < 0$; $\chi_t > 0$, $\psi_t > 0$ if $\lambda_t = 0$; and $\chi_t \geq 0$, $\psi_t > 0$ if $\lambda_t > 0$. For the GIG parameters, PP13 propose two specifications: (i) the iid case, where G_t are iid with

time-invariant parameters, i.e., $\lambda_t = \lambda$, $\chi_t = \chi$ and $\psi_t = \psi$; and (ii) the stochastic volatility case, where $G_t \mid \mathbf{\Phi}_{t-1}$ has time dependent parameters with the dynamics described by a system of conditional moment equations

$$\mathbb{E}\left[G_t^r \mid \mathbf{\Phi}_{t-1}\right] = c_r + \rho_r \mathbb{E}\left[G_{t-1}^r \mid \mathbf{\Phi}_{t-2}\right] + \sigma_r \zeta_{r,t}\,, \tag{4.4}$$

for a set of positive integer values of r; $\zeta_{r,t} = \mathbb{E}\left[G_t^r \mid \mathbf{\Phi}_t\right] - \mathbb{E}\left[G_t^r \mid \mathbf{\Phi}_{t-1}\right]$; and c_r, ρ_r and σ_r are parameters to be estimated. As discussed in PP13, $\zeta_{r,t}$ represents the unpredictable component affecting inference about the rth moment of the mixing variable G_t. It is a martingale difference sequence (MDS) with respect to $\mathbf{\Phi}_{t-1}$, thus $\mathbb{E}\left[\zeta_{r,t}\right] = 0$, $\mathrm{Cov}\left(\zeta_{r,t}, \zeta_{r,t-s}\right) = 0$, $s = 1, 2, \ldots$, and it can be used as a driver of the dynamics in (4.4). The model with formulation (4.4) is called a hybrid GARCH-Stochastic Volatility (GARCH-SV) extension. The link between this and the SV model of Taylor (1982) is detailed in PP13.

Due to the MDS property of the $\zeta_{r,t}$ innovations, the conditional forecasts of the future conditional moments of G_t are given by

$$\mathbb{E}\left[G_{t+s}^r \mid \mathbf{\Phi}_t\right] = c_r \sum_{i=0}^{s-1} \rho_r^i + \rho_r^s\, \mathbb{E}\left[G_t^r \mid \mathbf{\Phi}_{t-1}\right], \quad s \geq 1, \tag{4.5}$$

where $\mathbb{E}\left[G_t^r \mid \mathbf{\Phi}_{t-1}\right]$ is measurable with respect to the information up to time $t-1$ and is given by

$$\mathbb{E}\left[G_t^r \mid \mathbf{\Phi}_{t-1}\right] = \left(\frac{\chi_t}{\psi_t}\right)^{r/2} \frac{K_{\lambda_t + r}\left(\sqrt{\chi_t \psi_t}\right)}{K_{\lambda_t}\left(\sqrt{\chi_t \psi_t}\right)}, \quad r \in \mathbb{R}, \tag{4.6}$$

which involves a ratio of Bessel functions $K_v\left(z\right)$ as given in (4.3). If $|\rho_r| < 1$, then the process in (4.4) is mean-reverting, and for $s \to \infty$, the forecast approaches the unconditional mean value $c_r / \left(1 - \rho_r\right)$ of G_t^r.

The conditional, positive definite, dispersion matrix $\mathbf{H}_t$ is decomposed as

$$\mathbf{H}_t \equiv \mathbf{S}_t \mathbf{\Gamma}_t \mathbf{S}_t, \tag{4.7}$$

where $\mathbf{S}_t$ is a diagonal matrix composed of the strictly positive conditional scale terms $s_{k,t}$, $k = 1, \ldots, K$, and $\mathbf{\Gamma}_t = \mathrm{Corr}_{t-1}\left(\mathbf{Y}_t \mid G_t\right)$ is a dependency matrix. The univariate scale terms $s_{k,t}$ are modeled by a GARCH-type process, e.g., the GARCH(1,1) model

$$s_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{k,t-1}^2 + \beta_k s_{k,t-1}^2, \tag{4.8}$$

where $\varepsilon_{k,t} = y_{k,t} - \mu_k - \gamma_k G_t$ is the kth element of the $\boldsymbol{\varepsilon}_t$ vector in (4.1), and $\omega_k > 0$, $\alpha_k \geq 0$, $\beta_k \geq 0$, for $k = 1, 2, \ldots, K$. More general formulations could be used. In particular, we consider also the GJR-GARCH$(1,1)$ model of Glosten et al. (1993), which can capture an asymmetry effect, with the dynamics given by

$$s_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{k,t-1}^2 + \eta_k \varepsilon_{k,t-1}^2 \mathbf{1}_{\{\varepsilon_{k,t-1} > 0\}} + \beta_k s_{k,t-1}^2, \tag{4.9}$$

where $\mathbf{1}_{\{\}}$ is an indicator function, and $\eta_k \geq 0$ captures asymmetry in the scale-term response to the last period innovation.

The CCC version of the model, proposed in PP13, assumes that the dependency matrix is time invariant $\mathbf{\Gamma}_t = \mathbf{\Gamma}$. Here, we adopt the following structures to induce time-varying conditional correlation:

(i) Varying Correlation (VC) model of Tse and Tsui (2002)

$$\mathbf{\Gamma}_t = (1 - a - b)\,\mathbf{S} + a\mathbf{\Psi}_{t-1} + b\mathbf{\Gamma}_{t-1}, \tag{4.10}$$

where $\mathbf{\Psi}_{t-1}$ is a sample correlation matrix of the past $M \geq K$ ($M \geq K$ is a necessary condition for $\mathbf{\Psi}_{t-1}$ to be positive definite) standardized residuals $\boldsymbol{e}_{t-1}, \ldots, \boldsymbol{e}_{t-M}$, where $\boldsymbol{e}_{t-m} = G_{t-m}^{-1/2}\mathbf{S}_{t-m}^{-1}\boldsymbol{\varepsilon}_{t-m}$, for $m = 1, \ldots, M$, with the unobserved realizations of $G_{t-m}^{1/2}$ are replaced by their conditional expectations, $\mathbb{E}\left[G_t^{1/2} \mid \mathbf{\Phi}_t\right]$, from the E-step of the EM algorithm given below.

(ii) Dynamic Conditional Correlation (DCC) model of Engle (2002) with the $\mathbf{Q}_t$-dynamics

$$\mathbf{Q}_t = (1 - a - b)\,\mathbf{S} + aG_{t-1}^{-1}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + b\mathbf{Q}_{t-1}, \tag{4.11}$$

where G_t^{-1} are, similar to above, replaced by their conditional expectations, $\mathbb{E}\left[G_t^{-1} \mid \mathbf{\Phi}_t\right]$, from the E-step of the EM algorithm.

(iii) Corrected Dynamics Conditional Correlation (cDCC) model of Aielli (2011) with the $\mathbf{Q}_t$-dynamics governed by

$$\mathbf{Q}_t = (1 - a - b)\,\mathbf{S} + a\left(\mathbf{I}_K \odot \mathbf{Q}_{t-1}\right)^{1/2} G_{t-1}^{-1}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'\left(\mathbf{I}_K \odot \mathbf{Q}_{t-1}\right)^{1/2} + b\mathbf{Q}_{t-1}, \tag{4.12}$$

where $\mathbf{I}_K$ is a $K \times K$ identity matrix, $\odot$ is the Hadamard product (element by element multiplication), and G_t^{-1}, again are replaced by their conditional expectations, $\mathbb{E}\left[G_t^{-1} \mid \mathbf{\Phi}_t\right]$, from the E-step of the EM algorithm.

In case of (4.11) and (4.12), the dependency matrices, $\mathbf{\Gamma}_t$, for $t = 1, \ldots, T$ are obtained by re-scaling $\mathbf{Q}_t$ by

$$\mathbf{\Gamma}_t = \left(\mathbf{I}_K \odot \mathbf{Q}_t\right)^{-1/2}\mathbf{Q}_t\left(\mathbf{I}_K \odot \mathbf{Q}_t\right)^{-1/2}. \tag{4.13}$$

The dynamics given in (4.10), (4.11) and (4.12) are obvious analogs of the existing models, with the G_t factor introduced to accommodate the MGHyp distribution. As opposed to the Gaussian case, $\mathbf{\Gamma}_t$ are not conditional (on the past information) correlation matrices of the returns. They are dependency matrices, and, for all t, still required to be positive definite, symmetric, with ones on the main diagonal, and with off-diagonal elements between -1 and 1. Therefore, analogously to the Gaussian case, the sufficient conditions for $\mathbf{\Gamma}_t$ to be well-defined, for all models, for all $t > 0$ are $a \geq 0$, $b \geq 0$ and $a + b < 1$; and $\mathbf{Q}_0$, $\mathbf{\Gamma}_0$, and $\mathbf{S}$ have to be well-defined dependency matrices.

For notational convenience later, we collect the parameters of the model into three vectors (process, distribution, and correlation)

$$\begin{aligned} \boldsymbol{\theta}_P &= \left(\boldsymbol{\mu}', \boldsymbol{\gamma}', \boldsymbol{\omega}', \boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\eta}'\right)', \quad \boldsymbol{\theta}_D = \left(\mathbf{c}', \boldsymbol{\rho}', \boldsymbol{\sigma}'\right)', \text{ and} \\ \boldsymbol{\theta}_C &= \left(\operatorname{vech}\left(\mathbf{\Gamma}_0\right)', \operatorname{vech}\left(\mathbf{S}\right)', a, b\right)', \end{aligned} \tag{4.14}$$

where $\boldsymbol{\omega}, \boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are K-dimensional vectors of GJR-GARCH$(1,1)$ parameters from (4.9) (in case of GARCH$(1,1)$, from (4.8), $\boldsymbol{\eta} = \mathbf{0}$); $\mathbf{c}$, $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ denote vectors of c_r, ρ_r and σ_r parameters, respectively (but in case of constant G_t parameters, $\boldsymbol{\theta}_D$ reduces to (λ, χ, ψ)); vech $(\boldsymbol{\Gamma}_0)$ and vech $(\mathbf{S})'$ denote a column vector of the elements above the main diagonal of matrix $\boldsymbol{\Gamma}_0$ and $\mathbf{S}$, respectively, and for the CCC model, $\boldsymbol{\theta}_C$ reduces to vech $(\boldsymbol{\Gamma})$.

In the estimation procedure, the starting values are set as follows. Those of the asymmetry parameter, γ_k, and the GJR-GARCH parameter, η_k, are taken to be zero, and the remaining ones in $\boldsymbol{\theta}_P$, $(\mu_k, \omega_k, \alpha_k, \beta_k)$, to be those values obtained from the normal-based GARCH estimates, using the method in Paolella and Polak (2013a) to avoid inferior local likelihood maxima. For $\boldsymbol{\theta}_C$, we use $a = 0.05$, $b = 0.85$, and $\mathbf{Q}_0 = \boldsymbol{\Gamma}_0 = \mathbf{S}$, where for $\mathbf{S}$ we follow the correlation targeting approach (see, e.g., Engle, 2002) and set it equal to the unconditional covariance matrix of the standardized residuals, $\boldsymbol{e}_t$, with one exception of this rule for the cDCC model, for which, as Aielli (2011) advocates, we use $\mathbf{S} = \mathbb{E}[(\mathbf{I}_K \odot \mathbf{Q}_t)^{1/2} \boldsymbol{e}_t \boldsymbol{e}_t' (\mathbf{I}_K \odot \mathbf{Q}_t)^{1/2}]$. As in PP13, for $\boldsymbol{\theta}_D = (\lambda, \chi, \psi)$, we have confirmed that the likelihood is such that optimization is rather robust to the choice of starting values, e.g., for the special case of the Laplace distribution used in the empirical section below, we use $\lambda = 2$ as the starting value. For the hybrid GARCH-SV model we set $c = 0.1$ and $\rho = 0.8$. In the estimation with a rolling window, we use also as starting values the previous window estimates; and take the final estimates to be those with the higher likelihood value.

In the original VC, DCC and cDCC models (i.e., under the Gaussianity assumption) a one-step ahead forecast is straightforward, because, at time t, we know all the terms required to compute $\mathbf{Q}_{t+1}$. When we move to the MGHyp distribution and want to forecast the dependency matrix, $\boldsymbol{\Gamma}_{t+1}$, the situation is only slightly more complicated, because one needs to replace the unobserved G_t factor by its expectation conditional on all the information that is available at time t, i.e., $\mathbb{E}[G_t \mid \boldsymbol{\Phi}_t]$. In case of (4.11) and (4.12) dynamics, given the forecast of $\mathbf{Q}_{t+1}$, one applies the re-scaling from (4.13) to get the dependency matrix forecast $\widehat{\boldsymbol{\Gamma}}_{t+1}$. Forecasts more than one-step ahead are also possible in our models and are analogous to the corresponding forecasts under Gaussianity.

The conditional mean and covariance matrix of $\mathbf{Y}_t$, implied by the model (4.1), are given by

$$\mathbb{E}[\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1}] = \boldsymbol{\mu} + \mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}] \boldsymbol{\gamma} \tag{4.15}$$

and

$$\mathrm{Cov}(\mathbf{Y}_t \mid \boldsymbol{\Phi}_{t-1}) = \mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}] \mathbf{H}_t + \mathbb{V}(G_t \mid \boldsymbol{\Phi}_{t-1}) \boldsymbol{\gamma}\boldsymbol{\gamma}', \tag{4.16}$$

respectively, where $\mathbb{V}(G_t \mid \boldsymbol{\Phi}_{t-1}) = \mathbb{E}\left[G_t^2 \mid \boldsymbol{\Phi}_{t-1}\right] - (\mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}])^2$. It follows from (4.16) that the conditional correlation matrix has typical off-diagonal entry $\mathrm{c}_{i,j,t}$ given by

$$\mathrm{c}_{i,j,t} = \frac{\mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}] s_{i,t} s_{j,t} \Gamma_{i,j,t} + \mathbb{V}(G_t \mid \boldsymbol{\Phi}_{t-1}) \gamma_i \gamma_j}{\sqrt{\mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}] s_{i,t}^2 + \mathbb{V}(G_t \mid \boldsymbol{\Phi}_{t-1}) \gamma_i^2} \sqrt{\mathbb{E}[G_t \mid \boldsymbol{\Phi}_{t-1}] s_{j,t}^2 + \mathbb{V}(G_t \mid \boldsymbol{\Phi}_{t-1}) \gamma_j^2}}, \tag{4.17}$$

where $\Gamma_{i,j,t}$ is the element in row i and column j of the matrix $\boldsymbol{\Gamma}_t$.

The matrix $\boldsymbol{\Gamma}_t$ is a correlation matrix only conditionally on the realization of the mixing process. For this reason, we call $\boldsymbol{\Gamma}_t$ the dependency matrix. The dynamics in $\boldsymbol{\Gamma}_t$ imply that

the conditional correlation matrix of $\mathbf{Y}_t \mid \mathbf{\Phi}_{t-1}$ is time-varying. Additionally, in case of the hybrid GARCH-SV model, and for $\boldsymbol{\gamma} \neq \mathbf{0}$ and $\mathbb{E}[G_t \mid \mathbf{\Phi}_{t-1}] \neq \mathbb{V}(G_t \mid \mathbf{\Phi}_{t-1})$, the dynamics in the parameters of $G_t \mid \mathbf{\Phi}_{t-1}$ in (4.4), are a second source of the dynamics in the correlations. Empirical results given below confirm that both sources of the dynamics improve the forecasting performance of the model. All further conditional moments implied by the model (including the limiting cases of the mixing law) are available in Scott et al. (2011), while for the co-skewness and co-kurtosis matrices, see Mencía and Sentana (2009).

Finally, the volatility shock of the SV component influences (in a multiplicative way) each of the asset-specific conditional volatilities via (4.1b), and drives the dynamics of higher conditional moments (e.g., skewness and kurtosis) and co-moments of the returns. Because of the asset-specific conditional asymmetry coefficient γ_k, the impact of the SV component on each volatility is not equal across assets and its conditional expected value varies over time; see PP13 for a detailed discussion and an empirical demonstration of this.

4.3 Estimation

The estimation is conducted as in PP13, but augmented to accommodate the correlation dynamics, as detailed below. The algorithm consists of the E-step, in which the realizations of the unobserved mixing variables $\{G_t\}_{t=1,\dots,T}$ are imputed (in particular into the GARCH-type recursion); and the CM-steps, in which all the parameters are updated by maximizing either the unconditional likelihood function $L_{\mathbf{Y}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C)$, or the conditional one, $L_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C)$.

The steps of the algorithm, in case of our model, are based on the following complete log-likelihood function decomposition

$$\begin{aligned} \log L_{\mathbf{Y},\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C) &= \log L_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C) + \log L_{\mathbf{G}}(\boldsymbol{\theta}_D) \\ &= \log L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P) + \log L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C) + \log L_{\mathbf{G}}(\boldsymbol{\theta}_D), \end{aligned} \tag{4.18}$$

where

(i) $L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P)$ is the mean-volatility term given by

$$\begin{aligned} \log L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P) = -\frac{1}{2}\sum_{t=1}^{T}\Big[&K\log(2\pi) + \log|\mathbf{S}_t|^2 \\ &+ g_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma} g_t)'\mathbf{S}_t^{-1}\mathbf{S}_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma} g_t) + \log g_t\Big], \end{aligned} \tag{4.19}$$

(ii) $L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C)$ is the correlation term given by

$$\log L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log|\mathbf{\Gamma}_t| + \boldsymbol{e}_t'\mathbf{\Gamma}_t^{-1}\boldsymbol{e}_t - \boldsymbol{e}_t'\boldsymbol{e}_t\right], \tag{4.20}$$

where $\boldsymbol{e}_t = g_t^{-1/2}\mathbf{S}_t^{-1}\boldsymbol{\varepsilon}_t$, $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma} g_t$ from (4.1), and $\mathbf{\Gamma}_t$ have the dynamics described by (4.10) or (4.13) with (4.11) or (4.12);

(iii) $L_{\mathbf{G}}(\boldsymbol{\theta}_D)$ denotes the likelihood function of $(G_t \mid \boldsymbol{\Phi}_{t-1}) \sim \text{GIG}(\lambda_t, \chi_t, \psi_t)$, $t = 1, 2, \ldots, T$.

Given (4.18), estimates of $\boldsymbol{\theta}_P$, $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_D$ can be obtained by the following ECME algorithm.

E-step: Calculate $\mathbb{E}\big[\log L_{\mathbf{Y},\mathbf{G}} \mid \mathbf{Y}; \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_D, \widehat{\boldsymbol{\theta}}_C\big]$.

The log-likelihood function (4.19) is linear with respect to g_t and g_t^{-1}. Hence the E-step involves replacing unobserved realizations of G_t and G_t^{-1} in (4.18) by the imputed values, $\widehat{G}_t$. Calculation shows that (see, e.g., Paolella, 2013, Eq. 35)

$$\left(G_t \mid \boldsymbol{\Phi}_t; \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_C, \widehat{\boldsymbol{\theta}}_D\right) \sim \text{GIG}\left(\lambda_t^*, \chi_t^*, \psi_t^*\right), \tag{4.21}$$

where

$$\lambda_t^* = \lambda_t - K/2,\ \chi_t^* = \chi_t + (\mathbf{y}_t - \widehat{\boldsymbol{\mu}})' \mathbf{S}_t^{-1} \widehat{\boldsymbol{\Gamma}}_t^{-1} \mathbf{S}_t^{-1} (\mathbf{y}_t - \widehat{\boldsymbol{\mu}}) \text{ and } \psi_t^* = \psi_t + \widehat{\boldsymbol{\gamma}}' \mathbf{S}_t^{-1} \widehat{\boldsymbol{\Gamma}}_t^{-1} \mathbf{S}_t^{-1} \widehat{\boldsymbol{\gamma}}.$$

The latent values of g_t and g_t^{-1} are then updated by their conditional expectations from (4.21), using the expression for the moments of the GIG random variable given by (4.6) above.

CM1-step: Update $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_C$.

(P) Update $\boldsymbol{\theta}_P$ by computing

$$\arg\max_{\boldsymbol{\theta}_P} \log L^{\text{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P), \tag{4.22}$$

where $L^{\text{MV}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_P)$ is a Gaussian likelihood with zero correlation, so we can estimate the parameters of each asset, $(\mu_k, \gamma_k, \omega_k, \alpha_k, \beta_k)$, separately by maximizing the corresponding likelihood function.

(C) Update the DCC parameters $\boldsymbol{\theta}_C$ by computing

$$\arg\max_{\boldsymbol{\theta}_C} \log L^{\text{Corr}}_{\mathbf{Y}|\mathbf{G}}(\boldsymbol{\theta}_C), \tag{4.23}$$

where $L^{\text{Corr}}_{\mathbf{Y}|\mathbf{G}}\left(\widehat{\boldsymbol{\theta}}_P, \boldsymbol{\theta}_C\right)$ is given in (4.20), with the $\boldsymbol{e}_t$ replaced by the de-volatilized residuals $\widehat{G}_t^{-1/2} \mathbf{S}_t^{-1} \widehat{\boldsymbol{\varepsilon}}_t$, where $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\gamma}} \widehat{G}_t$, $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\gamma}}$ are obtained in part (P) directly above, and $\widehat{G}_t$ is obtained in the E-step.

CM2-step: Given the CM1-step estimates of $\boldsymbol{\theta}_P$, obtain new estimates of $\boldsymbol{\theta}_D$ by maximizing the incomplete data log-likelihood function, i.e., compute

$$\arg\max_{\boldsymbol{\theta}_D} \log L_{\mathbf{Y}}\left(\boldsymbol{\theta}_D \mid \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_C\right). \tag{4.24}$$

The above steps need to be iterated until convergence.

As argued in PP13, the MGHyp is too general, and we advocate use of special cases. In particular, and as detailed in PP13, we use the multivariate asymmetric Laplace (MALap), the multivariate normal inverse Gaussian (MNIG), and the multivariate asymmetric t-distribution (MAt). Each allows for individual asset asymmetry parameters, as well as higher kurtosis than the normal distribution.

4.4 Portfolio Construction and Risk Measures

Under the dynamics in (4.1), the conditional distribution of the multivariate set of asset returns is $\left(\mathbf{Y}_t \mid \mathbf{\Phi}_{t-1}\right) \sim \text{MGHyp}\left(\boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{H}_t, \lambda_t, \chi_t, \psi_t\right)$. The generalized hyperbolic class, including all the limiting cases, is closed under linear operations, so that the density of the portfolio $P_t = \mathbf{w}'\mathbf{Y}_t$ is given by

$$P_t \mid \mathbf{\Phi}_{t-1} \sim \text{GHyp}\left(\mu_{\mathbf{w}}, \gamma_{\mathbf{w}}, \sigma^2_{t,\mathbf{w}}, \lambda_t, \chi_t, \psi_t\right),$$

for some portfolio weight vector $\mathbf{w} = (w_1, w_2, \ldots, w_K)$ such that $\sum_{k=1}^K w_k = 1$, $\mu_{\mathbf{w}} = \mathbf{w}'\boldsymbol{\mu}$, $\gamma_{\mathbf{w}} = \mathbf{w}'\boldsymbol{\gamma}$, and $\sigma^2_{t,\mathbf{w}} = \mathbf{w}'\mathbf{H}_t\mathbf{w}$.

As an example, if we assume that the model innovations are MALap, then the conditional portfolio density $f_{P_t}(p; \mu_{\mathbf{w}}, \gamma_{\mathbf{w}}, \sigma_{t,\mathbf{w}}, \lambda_t)$ is ALap, with density given by

$$\frac{2\exp\left\{(p-\mu_{\mathbf{w}})\,\gamma_{\mathbf{w}}/\sigma^2_{t,\mathbf{w}}\right\}}{\sqrt{(2\pi)}\Gamma\left(\lambda_t\right)|\sigma_{t,\mathbf{w}}|}\left(\frac{m_t}{2+\gamma^2_{\mathbf{w}}/\sigma^2_{t,\mathbf{w}}}\right)^{\lambda_t/2-1/4} K_{\lambda_t-1/2}\left(\sqrt{m_t\left(2+\gamma^2_{\mathbf{w}}/\sigma^2_{t,\mathbf{w}}\right)}\right), \quad (4.25)$$

where $m_t = (p-\mu_{\mathbf{w}})^2/\sigma^2_{t,\mathbf{w}}$. This is plotted in Figure 4.1. In the symmetric case, the normal density with the same variance is plotted for comparison. Observe that, the higher the λ, the higher the standard deviation and the lower the kurtosis. In the center of the support, and for high values of λ, the Laplace density is approaching the normal density with the same variance, but eventually, further into the tails, as long as $\lambda < \infty$, the power factor of the tail of the Laplace density starts to dominate.

Given the conditional density of the portfolio f_{P_t}, the value at risk (VaR) and the expected shortfall (ES) can be computed. The former, given the tail probability α, is given by

$$\text{VaR}_\alpha^{t|t-1}\left(P_t\right) = -Q_\alpha\left(P_t \mid \Phi_{t-1}\right) = -\inf\left\{x \in \mathbb{R} : \mathbb{P}\left(P_t \le x \mid \mathbf{\Phi}_{t-1}\right) \ge \alpha\right\}, \quad (4.26)$$

where Q_α denotes the quantile function. The latter is

$$\text{ES}_\alpha^{t|t-1}\left(P_t\right) = \frac{1}{\alpha}\int_{-\infty}^{Q_\alpha(P_t|\Phi_{t-1})} p f_{P_t}\left(p \mid \mu_{\mathbf{w}}, \gamma_{\mathbf{w}}, \sigma_{t,\mathbf{w}}, \lambda_t, \chi_t, \psi_t\right) \mathrm{d}p, \quad (4.27)$$

where $\mathbf{H}_t$ in $\sigma^2_{t,\mathbf{w}}$ comes from the GARCH filter given the information in $\mathbf{\Phi}_{t-1}$, and, in case of the SV version of the model, $\lambda_t, \chi_t, \psi_t$ are obtained through (4.5) and (4.6).

As the GHyp cumulative distribution function (cdf), its inverse, and the tail conditional expectation are not available in explicit form, they have to be evaluated numerically. Once computed from the predictive portfolio distribution, one can conduct minimum ES portfolio optimization. Of course, mean-variance optimization can also be considered based on the quantities in (4.15) and (4.16), or, more generally, the mean-variance-skewness optimization as in Mencía and Sentana (2009).

Similar to Broda and Paolella (2011, 2010), which provide formulae for the ES for various distributions, including discrete mixtures, stable Paretian distributions, some special cases of (4.27), and a fast, accurate approximation in the noncentral Student's t case, we provide the following result, which reduces computation time.

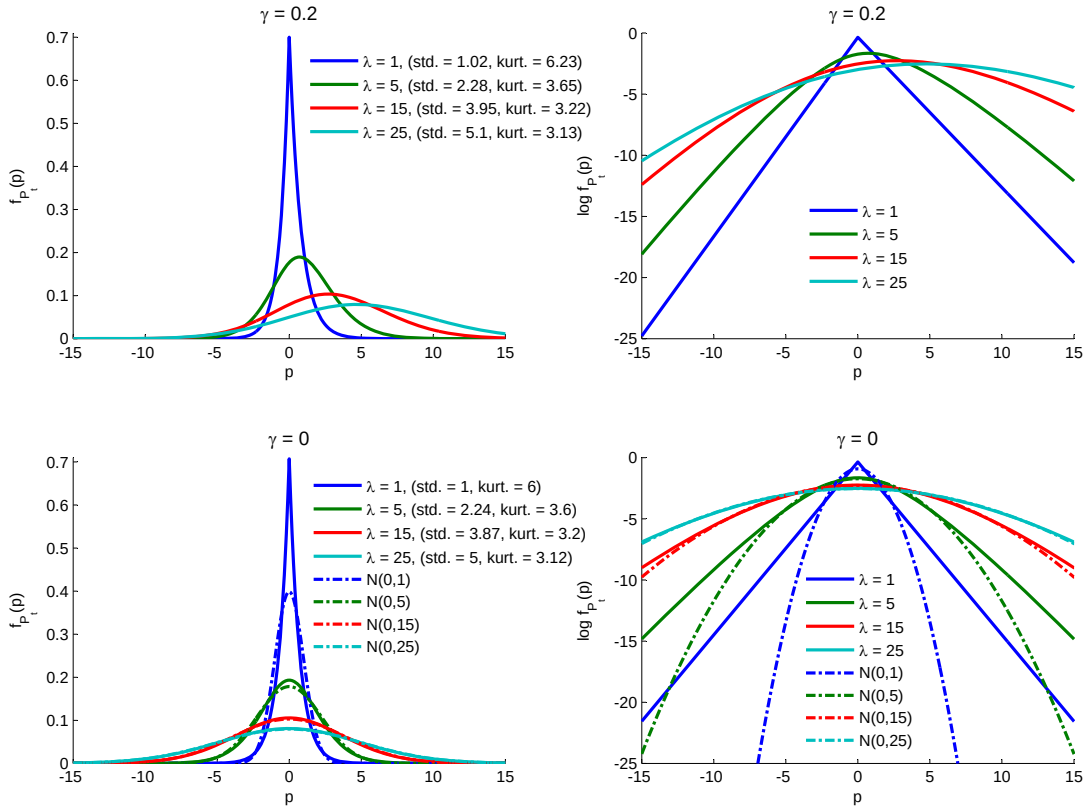

Figure 4.1: Univariate Laplace density plots with $\mu = 0$, $\sigma = 1$ for various γ and λ parameters. **Left column:** The density plots. Legends contain corresponding standard deviations and kurtosis coefficients computed based on the formulae in Scott et al. (2011). **Right column:** The log density plots. **First row:** Asymmetric case $\gamma = 0.2$. **Second row:** Symmetric case $\gamma = 0$. The dashed line corresponds to the normal density with the same variance as the corresponding Laplace law.

Proposition 4.4.1. *If* $P \sim \mathrm{GHyp}(\mu, \gamma, h, \lambda, \chi, \psi)$, *then*

$$\mathrm{ES}_\alpha(P) = -\mu - \frac{\gamma}{\alpha}\mathbb{E}[G] F_{P^*}(-\mathrm{VaR}_\alpha(P)) + \frac{h}{\alpha}\frac{C}{\sqrt{2\pi}}, \tag{4.28}$$

where F_{P^*} *is the cdf of* $P^* \sim \mathrm{GHyp}(\mu, \gamma, h, \lambda + 1, \chi, \psi)$*; and the constant* C *is given by*

$$C = \frac{\chi^{-\lambda}\left(\sqrt{\chi\psi}\right)^{\lambda}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} \frac{2K_{\widetilde{\lambda}}\left(\sqrt{\widetilde{\chi}\widetilde{\psi}}\right)}{\widetilde{\chi}^{-\widetilde{\lambda}}\left(\sqrt{\widetilde{\chi}\widetilde{\psi}}\right)^{\widetilde{\lambda}}} \exp\left(-\frac{(\mathrm{VaR}_\alpha(P) + \mu)\gamma}{h^2}\right),$$

with

$$\widetilde{\lambda} = \lambda + 1/2, \qquad \widetilde{\chi} = \chi + \frac{(\mathrm{VaR}_\alpha(P) + \mu)^2}{h^2} \quad \text{and} \quad \widetilde{\psi} = \psi + \frac{\gamma^2}{h^2}.$$

The proof of Proposition 4.4.1 is given in Appendix 4.6.1. In the special case of the NIG, the formula in Proposition 4.4.1 applies directly; for the two limiting cases of ALap distribution and At distribution, we have, respectively:

Remark 4.4.1. *If* $P \sim \mathrm{ALap}(\mu, \gamma, h, \lambda)$*, then*

$$\mathrm{ES}_\alpha(P) = -\mu - \frac{\gamma}{\alpha}\mathbb{E}[G] F_{P^*}(-\mathrm{VaR}_\alpha(P)) + \frac{h}{\alpha}\frac{C}{\sqrt{2\pi}}, \tag{4.29}$$

where F_{P^*} *is the cdf of* $P^* \sim \mathrm{ALap}(\mu, \gamma, h, \lambda+1)$*; and the constant* C *is given by*

$$C = \frac{1}{\Gamma(\lambda)} \frac{2K_{\widetilde{\lambda}}\left(\sqrt{\widetilde{\chi}\widetilde{\psi}}\right)}{\widetilde{\chi}^{-\widetilde{\lambda}}\left(\sqrt{\widetilde{\chi}\widetilde{\psi}}\right)^{\widetilde{\lambda}}} \exp\left(-\frac{(\mathrm{VaR}_\alpha(P)+\mu)\gamma}{h^2}\right),$$

for

$$\widetilde{\lambda} = \lambda + \frac{1}{2}, \qquad \widetilde{\chi} = \frac{(\mathrm{VaR}_\alpha(P)+\mu)^2}{h^2}, \quad \text{and} \quad \widetilde{\psi} = 2 + \frac{\gamma^2}{h^2}.$$

Remark 4.4.2. *If* $P \sim \mathrm{At}(\mu, \gamma, h, v)$*, then*

$$\mathrm{ES}_\alpha(P) = -\mu - \frac{\gamma}{\alpha}\mathbb{E}[G] F_{P^*}(-\mathrm{VaR}_\alpha(P)) + \frac{h}{\alpha}\frac{C}{\sqrt{2\pi}}, \tag{4.30}$$

where F_{P^*} *is the cdf of* $P^* \sim \mathrm{GHyp}(\mu, \gamma, h, v/2+1, v, \psi)$*,* $\psi = 0$*, the constant* C *is given by*

$$C = \frac{(v/2)^{v/2}}{\Gamma(v/2)} \frac{2K_{\widetilde{\lambda}}\left(\sqrt{\widetilde{\chi}\widetilde{\psi}}\right)}{\widetilde{\chi}^{-\widetilde{\lambda}}\left(\sqrt{\widetilde{\chi}\widetilde{\psi}}\right)^{\widetilde{\lambda}}} \exp\left(-\frac{(\mathrm{VaR}_\alpha(P)+\mu)\gamma}{h^2}\right),$$

for

$$\widetilde{\lambda} = \frac{v}{2} + \frac{1}{2}, \qquad \widetilde{\chi} = \frac{(\mathrm{VaR}_\alpha(P)+\mu)^2}{h^2} - v, \quad \text{and} \quad \widetilde{\psi} = \frac{\gamma^2}{h^2}.$$

The ES formulae in (4.27) and (4.28) require computation of the VaR value, which we do by numerical inversion of the portfolio cdf. As the portfolio distribution is GHyp, the corresponding portfolio cdf can be rewritten as

$$\Pr(P_t \le r) = \int_0^\infty \Phi\left(\frac{r - \mu_{\mathbf{w}} - \gamma_{\mathbf{w}} g}{\sigma_{t,\mathbf{w}}}\right) f_{G_t}(g; \lambda_t, \chi_t, \psi_t)\mathrm{d}g, \tag{4.31}$$

where Φ is the cdf of standard normal random variable. We found empirically that use of the latter representation can reduce the computation time required for the numerical inversion in the computation of the VaR values by a factor of four.

If, additionally, the returns are assumed to be elliptical, i.e., $\boldsymbol{\gamma} = \mathbf{0}$, then, in comparison to numerical integration in (4.27), use of (4.28) reduces the computational time required for calculating the ES by a factor of over 100. In the general case $\boldsymbol{\gamma} \neq \mathbf{0}$, both approaches are equally fast. These computations are necessary for risk evaluation of a specific portfolio, but

when interest centers on portfolio optimization, the necessity to evaluate these measures for each candidate portfolio during the optimization process becomes quite a bottleneck. Hence, we suggest an alternative approach, as discussed next.

The minimum-variance portfolio and the mean-variance portfolio are the solutions of

$$\min_{\mathbf{w}\in\mathfrak{W}} \sigma^2_{t,\mathbf{w}}, \tag{4.32}$$

where $\mathfrak{W} = \left\{\mathbf{w} \in \mathbb{R}^K : \sum_{k=1}^K w_k = 1,\ w_k \geq 0, k = 1, \ldots, K\right\}$, i.e., we impose the short-selling constraint, assume no risk-free asset, and, in case of the classical Markowitz (1952) mean-variance portfolio, one needs to include an additional constraint for the minimum portfolio return $\mu_{\mathbf{w}} + \gamma_{\mathbf{w}}\mathbb{E}\left[G_t \mid \mathbf{\Phi}_{t-1}\right] \geq \bar{p}$, where $\bar{p}$ is a constant for the target minimum return of the portfolio.

The portfolio variance is the simplest risk measure, and can be evaluated almost instantaneously for every candidate portfolio. The minimum-ES portfolio and mean-ES portfolio solve optimization problems analogous to (4.32) with the conditional variance $\sigma^2_{t,\mathbf{w}}$ replaced by the conditional ES given in (4.27). This is computationally much more intensive. However, the computational efficiency can be increased by using the results from Rockafeller and Uryasev (2000). First, note that (4.27) can be rewritten as

$$\mathrm{ES}_\alpha^{t|t-1}\left(P_t\right) = -\left\{\mathrm{VaR}_\alpha^{t|t-1}\left(P_t\right) + \frac{1}{\alpha}\int_{-\infty}^{\infty}\left[p + \mathrm{VaR}_\alpha^{t|t-1}\left(P_t\right)\right]^- f_{P_t}\left(p \mid \mu_{\mathbf{w}}, \gamma_{\mathbf{w}}, \sigma_{t,\mathbf{w}}, \lambda_t, \chi_t, \psi_t\right)\mathrm{d}p\right\},$$

where $[x]^- = -x$ if $x \leq 0$, and $[x]^- = 0$ if $x > 0$. Rockafeller and Uryasev (2000) introduce an auxiliary function

$$F_\alpha\left(x, \mathbf{w}\right) = x + \frac{1}{\alpha}\int_{-\infty}^{\infty}\left[p + x\right]^- f_{P_t}\left(p \mid \mu_{\mathbf{w}}, \gamma_{\mathbf{w}}, \sigma_{t,\mathbf{w}}, \lambda_t, \chi_t, \psi_t\right)\mathrm{d}p,$$

for which they show that

$$\mathrm{ES}_\alpha^{t|t-1}\left(P_t\right) = \min_x F_\alpha\left(x, \mathbf{w}\right) \text{ and } \mathrm{VaR}_\alpha^{t|t-1}\left(P_t\right) = \arg\min_x F_\alpha\left(x, \mathbf{w}\right), \tag{4.33}$$

and that F_α is continuously differentiable and convex as a function of $(x, \mathbf{w})$. Hence, the ES and VaR can be computed simultaneously (without determining the VaR first) as a unique minimum and a unique minimizer, respectively, of a convex minimization problem. Furthermore, as shown in Rockafeller and Uryasev (2000),

$$\min_{(x,\mathbf{w})\in\mathbb{R}\times\mathfrak{W}} F_\alpha\left(x, \mathbf{w}\right) = \min_{\mathbf{w}\in\mathfrak{W}} \mathrm{ES}_\alpha^{t|t-1}\left(P_t\right), \tag{4.34}$$

where the pair $(x^*, \mathbf{w}^*)$ achieves the first minimum if and only if $\mathbf{w}^*$ achieves the second minimum and $x^* = \mathrm{VaR}_\alpha^{t|t-1}\left(\mathbf{w}^{*\prime}\mathbf{Y}_t\right)$. The advantages of the first minimization in (4.34) are that (i) it completely avoids the calculation of the ES (and VaR) of the candidate portfolios; and (ii) as both F_α and the constraint set are convex, it is a convex programming problem. As such, it can be solved more efficiently and provides a global optimum.

4.5 Empirical Results

The novelty of model (4.1), relative to the model in PP13, is the dynamics in the dependency matrix. To investigate the empirical consequences of these dynamics, we use the data set consisting of the $2,767$ daily returns of $K = 30$ components of the Dow Jones Industrial Index (DJ-30) from January 2nd, 2001, to December 30th, 2011 (based on the DJ-30 composition as of June 8th, 2009; see Appendix 4.6.2 for details). Returns for each asset are computed as continuously compounded percentage returns, given by $y_{k,t} = 100 \log (p_{k,t}/p_{k,t-1})$, where $p_{k,t}$ is the price of asset k at time t. The empirical comparison is rather extensive; it employs models with both GARCH$(1,1)$ and GJR-GARCH$(1,1)$ univariate dynamics, and with all the aforementioned dependency matrix dynamics and distributions. All together, we analyse eight Gaussian-based models, one iid heavy-tailed model, ten models from PP13, and thirty new dynamic models.

We state here the results of Q-Q plots, estimated over the entire data set (with the set of all figures available upon request). They demonstrate the improved tail fit when moving from Gaussian-DCC to MGHyp-DCC (in full agreement with the corresponding CCC results, demonstrated in PP13), and also the lack of improvement when moving from CCC to DCC, under the Gaussian assumption. This latter result implies that incorporation of dynamics in the correlations does not improve the fit in the tails—this finding being fully in line with theory, because the Gaussian law is asymptotically independent in the tails (see, e.g., McNeil et al. 2005, Ch. 5.3). Finally, it will be seen below that the best model (in terms of density forecasting) among the CCC models is MALap-CCC GARCH$(1,1)$-SV. When comparing its Q-Q plot to that of the best DCC model, MNIG-cDCC GARCH$(1,1)$-SV, we see that, when accounting for non-Gaussianity, a slight improvement in the tail fit of the DCC model is visible. This result also agrees with the forecasting results discussed below.

We turn now to the role of the distribution for the correlation dynamics. Figure 4.2 compares the dynamics of conditional correlations implied by MN-cDCC GARCH$(1,1)$, MNIG-cDCC hybrid GARCH$(1,1)$-SV, MAt-DCC GARCH$(1,1)$ and MNIG-cDCC GARCH$(1,1)$ models. Different panels correspond to different pairs of the DJ-30 components. We show only results for four models out of 49 considered, and for eight pairs of correlations out of over 400 estimated by the models, but the conclusions hold in general. The proposed MGHyp models produce very similar correlations, while the Gaussian models have systematically lower and more varying correlation estimates than the MGHyp models, and with some additional spikes in the dynamics. To corroborate these results, when looking at the portfolio performance, we find that those in the Gaussian case exhibit higher transaction costs. Moreover, the spikes in the dynamics cause large, temporal, and most probably spurious, changes in the portfolio weights, implying potential losses. Importantly, all of the models (Gaussian and non-Gaussian) capture an increase in the correlations during the recent financial crises of 2008.

With respect to out-of-sample forecasting performance, we estimate all the models using a rolling window of $1,000$ observations, and all criteria are based on one-step ahead predictions. The density predictions are compared using the normalized sum of the realized predictive log-

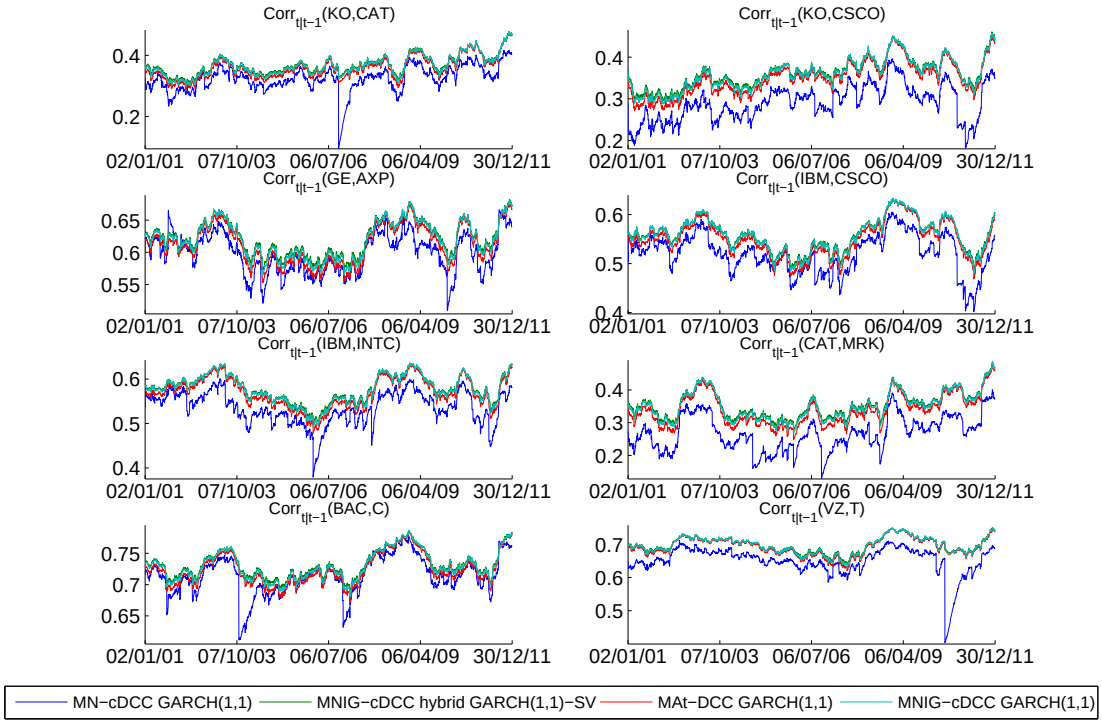

Figure 4.2: Dynamics of conditional correlations between returns of various pairs of assets (corresponding tickers are in the title of each panel) implied by different models (MN-cDCC GARCH$(1,1)$, NIG-cDCC GARCH$(1,1)$-SV, MAt-DCC GARCH$(1,1)$, and NIG-cDCC GARCH$(1,1)$).

likelihood values, which, for given model $\mathcal{M}$, is

$$S_T(\mathcal{M}) = \frac{1}{T}\sum_{t=1}^{T}\pi_t(\mathcal{M}), \qquad \text{where} \qquad \pi_t(\mathcal{M}) = \log f^{\mathcal{M}}_{t+1|t}(\mathbf{Y}_{t+1} \mid \boldsymbol{\theta}). \tag{4.35}$$

For the hybrid GARCH-SV models, a first-order approximation to π_t is used, and random parameters of the distribution of G_{t+1} are replaced with the values implied by the conditional expectation $\mathbb{E}[G_{t+1} \mid \boldsymbol{\Phi}_t]$.

The results are given in Table 4.1. The models are ordered according to the absolute distance of the failure rate for 1% VaR, from the nominal value of 0.01. The hybrid NIG-cDCC GARCH$(1,1)$-SV model performs best. It is closely followed by three other hybrid GARCH-SV models. Next in the ranking are the two MAt models. The Gaussian-based models perform the worst. Horizontal lines in the table draw attention to the MALap-IID model, which "separates" all the Gaussian-based models from the COMFORT models; this result being the same as found in PP13.

The second criteria is the accuracy of the VaR predictions implied by the proposed models. For this purpose, we compute daily one-day-ahead out-of-sample VaR forecasts at the 1%, 5%

$\mathcal{M}$	$S_T(\mathcal{M})$	$\mathcal{M}$	$S_T(\mathcal{M})$
NIG-cDCC GARCH(1, 1)-SV	−45.982	MALap-DCC GJR-GARCH(1, 1)-SV	−46.212
MALap-cDCC GARCH(1, 1)-SV	−46.001	NIG-DCC GJR-GARCH(1, 1)-SV	−46.230
MALap-DCC GARCH(1, 1)-SV	−46.009	MAt-VC GJR-GARCH(1, 1)	−46.235
NIG-DCC GARCH(1, 1)-SV	−46.009	NIG-cDCC GJR-GARCH(1, 1)	−46.236
MAt-cDCC GARCH(1, 1)	−46.011	MALap-VC GJR-GARCH(1, 1)-SV	−46.243
MAt-DCC GARCH(1, 1)	−46.016	MALap-cDCC GJR-GARCH(1, 1)	−46.262
MALap-VC GARCH(1, 1)-SV	−46.027	MALap-DCC GJR-GARCH(1, 1)	−46.282
NIG-cDCC GARCH(1, 1)	−46.036	NIG-DCC GJR-GARCH(1, 1)	−46.292
MAt-VC GARCH(1, 1)	−46.044	MALap-CCC GJR-GARCH(1, 1)-SV	−46.300
MALap-CCC GARCH(1, 1)-SV	−46.064	MAt-CCC GJR-GARCH(1, 1)	−46.301
NIG-CCC GARCH(1, 1)-SV	−46.064	MALap-VC GJR-GARCH(1, 1)	−46.323
NIG-DCC GARCH(1, 1)	−46.066	NIG-VC GJR-GARCH(1, 1)	−46.323
MALap-cDCC GARCH(1, 1)	−46.083	NIG-CCC GJR-GARCH(1, 1)-SV	−46.336
MALap-DCC GARCH(1, 1)	−46.094	MALap-CCC GJR-GARCH(1, 1)	−46.378
MAt-CCC GARCH(1, 1)	−46.097	NIG-CCC GJR-GARCH(1, 1)	−46.399
MALap-VC GARCH(1, 1)	−46.118	MALap-IID	−47.067
NIG-CCC GARCH(1, 1)	−46.133	MN-cDCC GARCH(1, 1)	−47.859
NIG-VC GARCH(1, 1)	−46.134	MN-DCC GARCH(1, 1)	−47.863
MALap-CCC GARCH(1, 1)	−46.161	MN-VC GARCH(1, 1)	−47.885
NIG-VC GARCH(1, 1)-SV	−46.190	MN-cDCC GJR-GARCH(1, 1)	−47.892
NIG-cDCC GJR-GARCH(1, 1)-SV	−46.196	MN-DCC GJR-GARCH(1, 1)	−47.895
NIG-VC GJR-GARCH(1, 1)-SV	−46.198	MN-VC GJR-GARCH(1, 1)	−47.914
MAt-cDCC GJR-GARCH(1, 1)	−46.200	MN-CCC GARCH(1, 1)	−47.951
MAt-DCC GJR-GARCH(1, 1)	−46.202	MN-CCC GJR-GARCH(1, 1)	−47.973
MALap-cDCC GJR-GARCH(1, 1)-SV	−46.204	MN-IID	−52.528

Table 4.1: Performance of the one-step ahead predictions of the return vector density for different models, $\mathcal{M}$, and measured by $S_T(\mathcal{M})$, in (4.35).

and 10% levels, for an equally weighted portfolio. The computation utilizes the results given in Section 4.4. At the nominal 1%, 5%, and 10% levels, the empirical coverage levels of the VaR forecasts obtained for various models are reported in Table 4.2. While there is no complete best performer, it is noteworthy that, for the 1% VaR results, the NIG distribution is the best, independently of the assumed dynamics, and hybrid GARCH-SV dynamics tend to improve the 1% VaR forecasts, supporting the fact that the SV dynamics in the G_t parameters are capturing dynamics in the tail behavior. Moreover, all the MGHyp models outperform the Gaussian based models, and the MALap-IID model again separates the two groups.

Figure 4.3 compares the density prediction results with the VaR backtesting results. All the MGHyp models provide better density forecasts than the Gaussian models, and also lead to better VaR predictions. The relation is not completely monotonic, but note that the empirical violation frequency (or, in short, failure rate) is not a unique criteria to evaluate the backtesting results. As discussed next, the independence of violations across time is also an important feature.

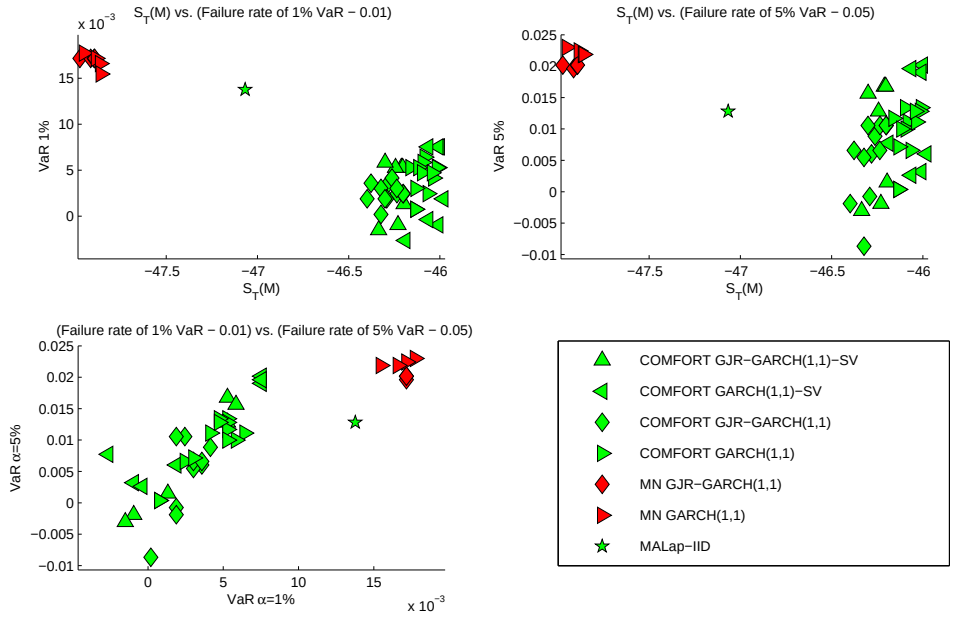

Figure 4.3: Cumulative predictive log-likelihood of various models vs. their VaR backtesting performance. **Upper left panel:** S_T vs. the dispersion from the optimal value of the 1% VaR failure rate. **Upper right panel:** S_T vs. the dispersion from the optimal value of the 5% VaR failure rate. **Bottom left panel:** The dispersion from the optimal value of the 1% VaR failure rate. vs. the dispersion from the optimal value of the 5% VaR failure rate.

Figure 4.4 illustrates 1% and 5% VaR forecasts across time, for an equally weighted portfolio of DJ-30 stocks, using three models—the best Gaussian model (MN-cDCC GARCH$(1,1)$), the "group separation" MALap-IID, and the best 1% VaR model (NIG-CCC GARCH$(1,1)$-SV).

$\mathcal{M}$	VaR 1%	VaR 5%	VaR 10%
NIG-CCC GARCH(1, 1)-SV	**0.0102**	0.0413	0.0838
NIG-VC GJR-GARCH(1, 1)	**0.0096**	0.0526	0.0905
NIG-VC GARCH(1, 1)	**0.0108**	**0.0504**	0.0872
NIG-DCC GJR-GARCH(1, 1)-SV	0.0091	0.0481	0.0849
NIG-DCC GARCH(1, 1)-SV	0.0091	0.0532	0.0900
NIG-cDCC GJR-GARCH(1, 1)-SV	0.0113	**0.0515**	0.0905
NIG-CCC GJR-GARCH(1, 1)-SV	0.0085	0.0470	0.0838
NIG-DCC GJR-GARCH(1, 1)	0.0119	**0.0492**	0.0911
NIG-CCC GJR-GARCH(1, 1)	0.0119	0.0481	0.0889
NIG-cDCC GARCH(1, 1)-SV	0.0119	0.0560	0.0951
MAt-CCC GJR-GARCH(1, 1)	0.0119	0.0606	0.0985
NIG-DCC GARCH(1, 1)	0.0125	0.0566	0.0962
MAt-VC GJR-GARCH(1, 1)	0.0125	0.0606	0.0979
MAt-cDCC GJR-GARCH(1, 1)	0.0125	0.0606	0.0979
MAt-DCC GJR-GARCH(1, 1)	0.0125	0.0606	0.0979
NIG-VC GARCH(1, 1)-SV	0.0074	0.0577	0.1138
MALap-VC GJR-GARCH(1, 1)	0.0130	0.0555	0.0934
NIG-cDCC GJR-GARCH(1, 1)	0.0130	0.0566	0.0951
NIG-CCC GARCH(1, 1)	0.0130	0.0572	0.0962
MALap-DCC GJR-GARCH(1, 1)	0.0136	0.0560	0.0951
MALap-CCC GJR-GARCH(1, 1)	0.0136	0.0566	0.0945
MALap-cDCC GJR-GARCH(1, 1)	0.0141	0.0589	0.0962
NIG-cDCC GARCH(1, 1)	0.0141	0.0611	**0.0996**
MAt-VC GARCH(1, 1)	0.0147	0.0628	0.1058
MAt-CCC GARCH(1, 1)	0.0147	0.0634	0.1058
MALap-VC GARCH(1, 1)	0.0153	0.0600	**0.1007**
MALap-CCC GARCH(1, 1)	0.0153	0.0617	**0.1002**
MALap-VC GJR-GARCH(1, 1)-SV	0.0153	0.0628	0.1013
MAt-DCC GARCH(1, 1)	0.0153	0.0628	0.1053
MAt-cDCC GARCH(1, 1)	0.0153	0.0634	0.1053
MALap-DCC GJR-GARCH(1, 1)-SV	0.0153	0.0668	0.1030
MALap-cDCC GJR-GARCH(1, 1)-SV	0.0153	0.0668	0.1041
MALap-DCC GARCH(1, 1)	0.0158	0.0600	0.1013
MALap-CCC GJR-GARCH(1, 1)-SV	0.0158	0.0656	0.1041
MALap-cDCC GARCH(1, 1)	0.0164	0.0611	0.1019
MALap-VC GARCH(1, 1)-SV	0.0175	0.0662	0.1053
MALap-DCC GARCH(1, 1)-SV	0.0175	0.0690	0.1058
MALap-CCC GARCH(1, 1)-SV	0.0175	0.0696	0.1053
MALap-cDCC GARCH(1, 1)-SV	0.0175	0.0702	0.1064
MALap-IID	0.0238	0.0628	**0.1007**
MN-cDCC GARCH(1, 1)	0.0255	0.0719	0.1041
MN-DCC GARCH(1, 1)	0.0266	0.0719	0.1041
MN-VC GJR-GARCH(1, 1)	0.0272	0.0696	0.1013
MN-DCC GJR-GARCH(1, 1)	0.0272	0.0702	**0.1007**
MN-cDCC GJR-GARCH(1, 1)	0.0272	0.0702	**0.1007**
MN-CCC GJR-GARCH(1, 1)	0.0272	0.0702	0.1013
MN-VC GARCH(1, 1)	0.0272	0.0724	0.1041
MN-CCC GARCH(1, 1)	0.0277	0.0730	0.1041

Table 4.2: One-day-ahead 1%, 5%, and 10% VaR forecasts failure rates for different models for an equally weighted portfolio of DJ-30 stocks, using various models. The models are ordered according to absolute distance of the failure rate for 1% VaR from 0.01, and in each category the best three results are in bold.

The Gaussian model has clearly too many violations, as confirmed in Table 4.2, while the MALap-IID model has a much better failure rate. However, from Figure 4.4, we see that almost all violations for the MALap-IID take place during the 2008 crisis, and hence they are strongly dependent. Moreover, outside of the crisis period, the predicted VaR values are much higher than those implied by the conditional models. Thus, during calm periods, the MALap-IID model predicts unnecessarily high capital requirements, and, during market downturns, it predicts insufficient capital holdings. These results are in line with other broad studies comparing iid and various conditional GARCH-type models, e.g., Mittnik and Paolella (2000) and Kuester et al. (2006a).

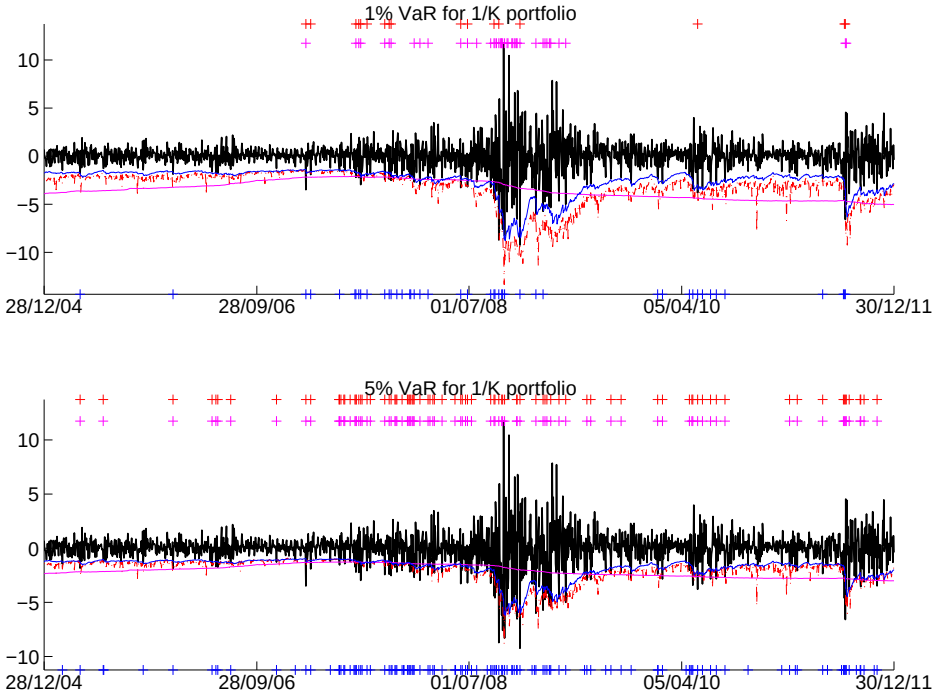

Figure 4.4: Returns, one-day-ahead 1% (5%) VaR forecasts, and VaR violations, for an equally weighted portfolio of DJ-30 stocks, using NIG-CCC GARCH(1, 1)-SV (dashes), MALap-IID (solid, magenta), and MN-cDCC GARCH(1, 1) (solid, blue) models. The VaR violations are depicted by + signs on top and in the bottom of the figures (with the same colors as VaR predictions).

Next we investigate the portfolio optimization results for different models. Table 4.3 reports the minimum-variance and minimum-ES portfolio performance for various models in terms of the annualized Sharpe ratio. The best models are from the MGHyp class, particularly with hybrid GARCH(1, 1)-SV dynamics. Interestingly, the GJR-GARCH(1, 1) models do not lead to better portfolio results. Together with all the Gaussian conditional models and the MALap-IID

model, they are (somewhat surprisingly) outperformed by the simplest MN-IID model. Models which lead to better density forecasts do not necessarily lead to better portfolios. However, as illustrated in Figure 4.5, all the MGHyp models outperform their counterpart Gaussian-based models. The bottom left panel in Figure 4.5 compares the minimum variance portfolios with minimum ES portfolios. The reason we do this comparison is that under the assumption of ellipticity, the two portfolio optimizations should lead to the same portfolios (and simulations confirm this). The MGHyp models account for non-ellipticity via the asymmetry vector $\boldsymbol{\gamma}$, and indeed most of the Sharpe ratios of the MGHyp models lie above the 45 degrees line, hence supporting the claim of non-ellipticity.

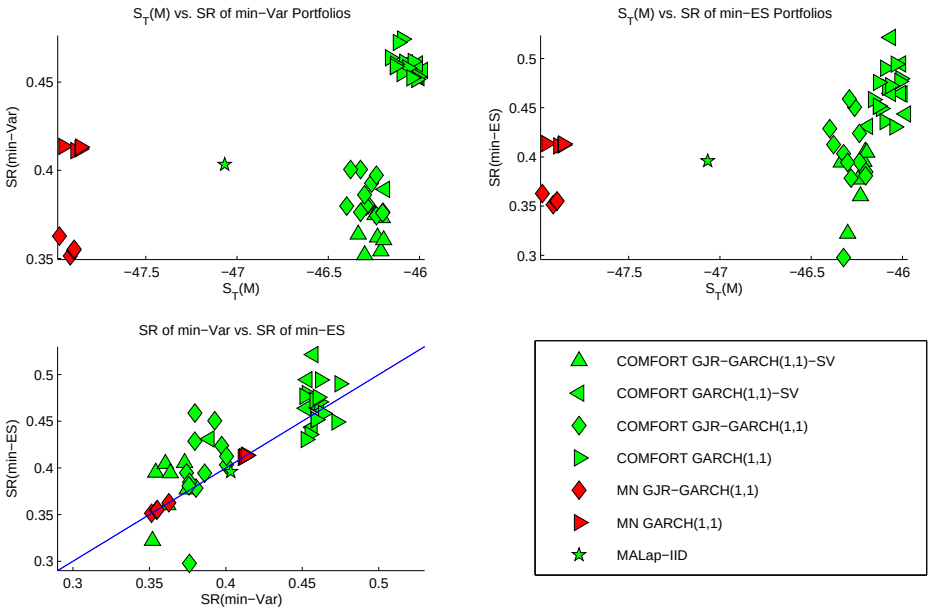

Figure 4.5: Cumulative predictive log-likelihood of various models vs. their portfolio performance. **Upper left panel:** S_T vs. Sharpe ratio of minimum variance portfolios. **Upper right panel:** S_T vs. Sharpe ratio of minimum expected shortfall portfolio. **Bottom left panel:** Sharpe ration of minimum variance portfolio vs. Sharpe ratio of minimum expected shortfall portfolio. (The diagonal line indicates models with equal performance, such as elliptical models.)

If interest centers not on minimum risk portfolios but the mean-variance or mean-ES portfolios, then the MGHyp models improve the portfolio performance in terms of average portfolio returns and portfolio volatilities, and hence leading to higher Sharpe ratios (see Figure 4.6). Finally, Figure 4.7 compares the models in terms of portfolio performance relative to risk backtesting performance (the latter being done for $1/K$ portfolio). The plots confirm that the MGHyp models outperform the Gaussian models in both categories.

$\mathcal{M}$	SR of min-Var	$\mathcal{M}$	SR of min-ES
MALap-DCC GARCH(1, 1)	0.4744	MALap-CCC GARCH(1, 1)-SV	0.5214
MALap-VC GARCH(1, 1)	0.4724	MALap-DCC GARCH(1, 1)-SV	0.4945
MALap-CCC GARCH(1, 1)	0.4637	NIG-cDCC GARCH(1, 1)	0.4942
NIG-cDCC GARCH(1, 1)	0.4617	MALap-DCC GARCH(1, 1)	0.4900
MALap-cDCC GARCH(1, 1)	0.4612	MAt-cDCC GARCH(1, 1)	0.4795
NIG-CCC GARCH(1, 1)-SV	0.4609	MAt-DCC GARCH(1, 1)	0.4768
NIG-DCC GARCH(1, 1)-SV	0.4603	NIG-CCC GARCH(1, 1)	0.4756
NIG-CCC GARCH(1, 1)	0.4602	NIG-DCC GARCH(1, 1)	0.4723
NIG-VC GARCH(1, 1)	0.4588	MALap-cDCC GARCH(1, 1)	0.4707
NIG-DCC GARCH(1, 1)	0.4581	MALap-VC GARCH(1, 1)-SV	0.4698
MALap-CCC GARCH(1, 1)-SV	0.4572	NIG-CCC GARCH(1, 1)-SV	0.4646
NIG-cDCC GARCH(1, 1)-SV	0.4566	NIG-DCC GARCH(1, 1)-SV	0.4644
MALap-VC GARCH(1, 1)-SV	0.4565	MALap-cDCC GARCH(1, 1)-SV	0.4640
MAt-CCC GARCH(1, 1)	0.4549	NIG-DCC GJR-GARCH(1, 1)	0.4588
MALap-DCC GARCH(1, 1)-SV	0.4532	MALap-CCC GARCH(1, 1)	0.4583
MAt-cDCC GARCH(1, 1)	0.4530	NIG-VC GARCH(1, 1)	0.4516
MALap-cDCC GARCH(1, 1)-SV	0.4525	MALap-cDCC GJR-GARCH(1, 1)	0.4505
MAt-VC GARCH(1, 1)	0.4524	MALap-VC GARCH(1, 1)	0.4492
MAt-DCC GARCH(1, 1)	0.4515	NIG-cDCC GARCH(1, 1)-SV	0.4436
MN-IID	0.4335	MAt-CCC GARCH(1, 1)	0.4358
MN-CCC GARCH(1, 1)	0.4136	MN-IID	0.4335
MN-cDCC GARCH(1, 1)	0.4132	NIG-VC GARCH(1, 1)-SV	0.4311
MN-DCC GARCH(1, 1)	0.4121	MAt-VC GARCH(1, 1)	0.4305
MN -VC GARCH(1, 1)	0.4111	NIG-CCC GJR-GARCH(1, 1)	0.4287
MALap-IID	0.4032	NIG-cDCC GJR-GARCH(1, 1)	0.4239
MALap-CCC GJR-GARCH(1, 1)	0.4005	MN-CCC GARCH(1, 1)	0.4136
MALap-VC GJR-GARCH(1, 1)	0.4004	MN-cDCC GARCH(1, 1)	0.4132
NIG-cDCC GJR-GARCH(1, 1)	0.3972	MALap-CCC GJR-GARCH(1, 1)	0.4125
MALap-cDCC GJR-GARCH(1, 1)	0.3928	MN-DCC GARCH(1, 1)	0.4121
NIG-VC GARCH(1, 1)-SV	0.3892	MN -VC GARCH(1, 1)	0.4111
MAt-CCC GJR-GARCH(1, 1)	0.3861	MALap-cDCC GJR-GARCH(1, 1)-SV	0.4055
MALap-DCC GJR-GARCH(1, 1)	0.3805	NIG-cDCC GJR-GARCH(1, 1)-SV	0.4046
NIG-DCC GJR-GARCH(1, 1)	0.3798	MALap-VC GJR-GARCH(1, 1)	0.4033
NIG-CCC GJR-GARCH(1, 1)	0.3797	MALap-IID	0.3959
MAt-cDCC GJR-GARCH(1, 1)	0.3763	MALap-DCC GJR-GARCH(1, 1)-SV	0.3952
NIG-VC GJR-GARCH(1, 1)	0.3762	MAt-VC GJR-GARCH(1, 1)	0.3948
MAt-DCC GJR-GARCH(1, 1)	0.3759	NIG-CCC GJR-GARCH(1, 1)-SV	0.3945
MALap-VC GJR-GARCH(1, 1)-SV	0.3747	MAt-CCC GJR-GARCH(1, 1)	0.3943
MAt-VC GJR-GARCH(1, 1)	0.3743	MAt-cDCC GJR-GARCH(1, 1)	0.3847
MALap-cDCC GJR-GARCH(1, 1)-SV	0.3730	MAt-DCC GJR-GARCH(1, 1)	0.3808
NIG-CCC GJR-GARCH(1, 1)-SV	0.3637	MALap-DCC GJR-GARCH(1, 1)	0.3784
MN-CCC GJR-GARCH(1, 1)	0.3627	MALap-VC GJR-GARCH(1, 1)-SV	0.3772
NIG-DCC GJR-GARCH(1, 1)-SV	0.3621	MN-CCC GJR-GARCH(1, 1)	0.3628
NIG-cDCC GJR-GARCH(1, 1)-SV	0.3605	NIG-DCC GJR-GARCH(1, 1)-SV	0.3603
MN-cDCC GJR-GARCH(1, 1)	0.3552	MN-cDCC GJR-GARCH(1, 1)	0.3549
MN-DCC GJR-GARCH(1, 1)	0.3547	MN-DCC GJR-GARCH(1, 1)	0.3544
MALap-DCC GJR-GARCH(1, 1)-SV	0.3542	MN -VC GJR-GARCH(1, 1)	0.3514
MALap-CCC GJR-GARCH(1, 1)-SV	0.3520	MALap-CCC GJR-GARCH(1, 1)-SV	0.3222
MN -VC GJR-GARCH(1, 1)	0.3514	NIG-VC GJR-GARCH(1, 1)	0.2977

Table 4.3: Performance of the minimum variance portfolio strategy (min-Var) without short-selling, and minimum expected shortfall portfolio strategy (min-ES) without short-selling, for different models, $\mathcal{M}$, and measured in terms of annualized Sharpe ratio (SR). For comparison Sharpe ratio of equally weighted portfolio $1/K$ is 0.1583.

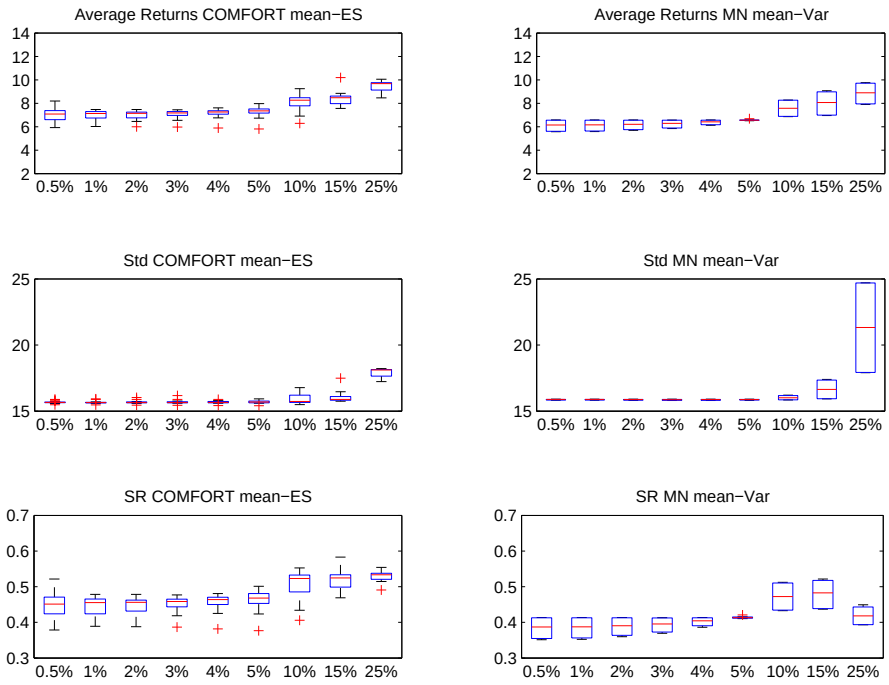

Figure 4.6: Boxplots of annualized portfolio returns, annualized portfolio standard deviations, and annualized Sharpe ratios, respectively, for different models and for different target minimum returns $\bar{p} \in \{0.5\%, 1\%, 2\%, 3\%, 4\%, 5\%, 10\%, 15\%, 25\%\}$. **Left panels:** The COMFORT models. **Right panels:** Gaussian conditional models.

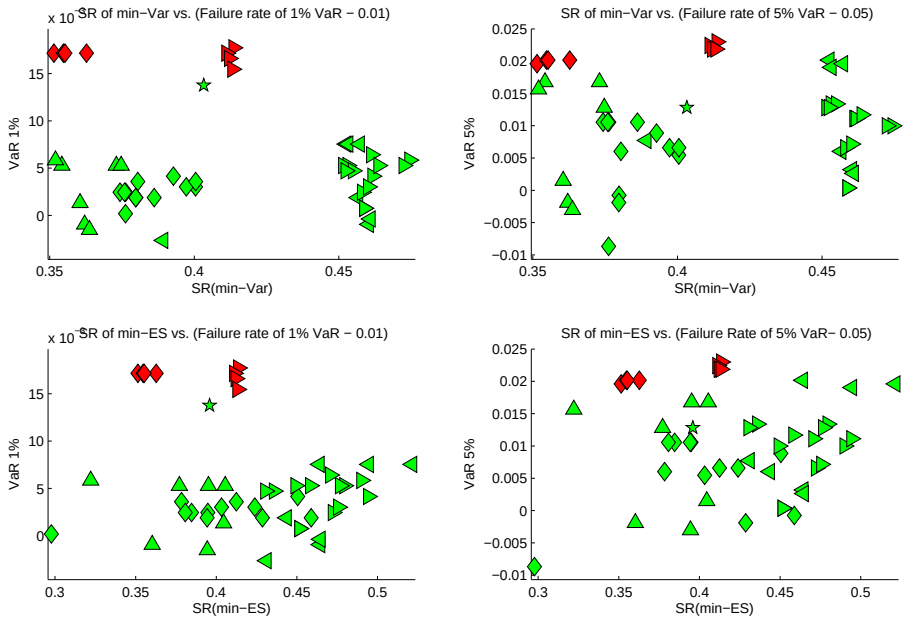

Figure 4.7: Portfolio performance of various models vs. their VaR backtesting performance. **Upper left panel:** Sharpe ratio of minimum variance portfolio vs. the dispersion from the optimal value of the 1% VaR failure rate. **Upper right panel:** Sharpe ratio of minimum variance portfolio vs. the dispersion from the optimal value of the 5% VaR failure rate. **Bottom left panel:** Sharpe ratio of minimum expected shortfall portfolio vs. the dispersion from the optimal value of the 1% VaR failure rate. **Bottom right panel:** Sharpe ratio of minimum expected shortfall portfolio vs. the dispersion from the optimal value of the 5% VaR failure rate. (The legend is the same as in Figure 4.5.)

4.6 Appendices

4.6.1 Proof of Proposition 4.4.1

Define the following integral for $n = 0, 1, \ldots, N$,

$$\mathbf{I}\left(q_{P,\alpha}, n; \mu, \gamma, h\right) = \int_{-\infty}^{q_{P,\alpha}} y^n f_P\left(y\right) \mathrm{d}y. \tag{4.36}$$

By (4.26) and (4.27), $\alpha = \mathbf{I}\left(-\mathrm{VaR}_\alpha\left(P\right), 0; \mu, \gamma, h\right)$ and $\mathrm{ES}_\alpha\left(P\right) = -\mathbf{I}\left(-\mathrm{VaR}_\alpha\left(P\right), 1; \mu, \gamma, h\right)/\alpha$, respectively. We start with a general result regarding $\mathbf{I}\left(q_{P,\alpha}, n; \mu, \gamma, h\right)$ for a mean-variance mixture of normal distribution.

Let $z = \left(y - \mu - \gamma g\right) / \left(h\sqrt{g}\right)$ and $c\left(g\right) = \left(q_{P,\alpha} - \mu - \gamma g\right) / \left(h\sqrt{g}\right)$. Then, changing variables,

$$\begin{aligned}
&\mathbf{I}\left(q_{P,\alpha}, n; \mu, \gamma, h\right) = \int_{-\infty}^{q_{P,\alpha}} y^n f_P\left(y\right) \mathrm{d}y = \int_{-\infty}^{q_{P,\alpha}} \int_0^\infty y^n f_{P|G}\left(y \mid g\right) f_G(g) \mathrm{d}g \mathrm{d}y \\
&= \int_0^\infty f_G\left(g\right) \int_{-\infty}^{q_{P,\alpha}} y^n f_{P|G}\left(y \mid g\right) \mathrm{d}y \mathrm{d}g \\
&= \int_0^\infty f_G\left(g\right) \int_{-\infty}^{q_{P,\alpha}} y^n \left(h\sqrt{g}\right)^{-1} f_{P|G}\left(\frac{y - \mu - \gamma g}{h\sqrt{g}} \mid g\right) \mathrm{d}y \mathrm{d}g \\
&= \int_0^\infty f_G\left(g\right) \int_{-\infty}^{c(g)} \left(h\sqrt{g}z + \mu + \gamma g\right)^n f_{Z|G}\left(z \mid g\right) \mathrm{d}z \mathrm{d}g
\end{aligned} \tag{4.37}$$

So, for $n = 0$ and $q_{P,\alpha} = -\mathrm{VaR}_\alpha\left(P\right)$, we have

$$\alpha = \mathbf{I}\left(-\mathrm{VaR}_\alpha\left(P\right), 0; \mu, \gamma, h\right) = \int_0^\infty f_G\left(g\right) \int_{-\infty}^{c(g)} f_{Z|G}\left(z \mid g\right) \mathrm{d}z \mathrm{d}g, \tag{4.38}$$

and, for $n = 1$ and $q_{P,\alpha} = -\mathrm{VaR}_\alpha\left(P\right)$, we have

$$\begin{aligned}
&\mathrm{ES}_\alpha\left(P\right) = -\frac{1}{\alpha}\mathbf{I}\left(-\mathrm{VaR}_\alpha\left(P\right), 1; \mu, \gamma, h\right) \\
&= -\frac{1}{\alpha}\int_0^\infty f_G\left(g\right) \int_{-\infty}^{c(g)} \left(h\sqrt{g}z + \mu + \gamma g\right) f_{Z|G}\left(z \mid g\right) \mathrm{d}z \mathrm{d}g \\
&= -\frac{\mu}{\alpha}\int_0^\infty f_G\left(g\right) \int_{-\infty}^{c(g)} f_{Z|G}\left(z \mid g\right) \mathrm{d}z \mathrm{d}g - \frac{\gamma}{\alpha}\int_0^\infty f_G\left(g\right) g \int_{-\infty}^{c(g)} f_{Z|G}\left(z \mid g\right) \mathrm{d}z \mathrm{d}g \\
&- \frac{h}{\alpha}\int_0^\infty f_G(g)\sqrt{g} \int_{-\infty}^{c(g)} z f_{Z|G}\left(z \mid g\right) \mathrm{d}z \mathrm{d}g.
\end{aligned} \tag{4.39}$$

From (4.38), the first integral in (4.39) is equal to α. The second integral in (4.39) is

$$\int_0^\infty f_G\left(g\right) g \int_{-\infty}^{c(g)} f_{Z|G}\left(z \mid g\right) \mathrm{d}z \mathrm{d}g = \mathbb{E}\left[G\right] F_{P^*}\left(-\mathrm{VaR}_\alpha\left(P\right)\right), \tag{4.40}$$

where F_{P^*} is the cumulative distribution function of $P^* \sim \mathrm{GHyp}\left(\mu, \gamma, h, \lambda + 1, \chi, \psi\right)$.

For the third integral in (4.39), note that, for any c,

$$\int_{-\infty}^{c} z f_{Z|G}\left(z \mid g\right) \mathrm{d}z = -\frac{1}{\sqrt{2\pi}}\Gamma\left(1, \frac{c^2}{2}\right),$$

where Γ is the upper incomplete gamma function which satisfies $\Gamma(1,x) = \exp(-x)$, for any $x > 0$; and, by the definition of $c(g)$,

$$-\frac{c^2(g)}{2} = -\frac{1}{2}\left(\frac{(\mathrm{VaR}_\alpha(P)+\mu)^2}{h^2}g^{-1} + \frac{\gamma^2}{h^2}g + \frac{2(\mathrm{VaR}_\alpha(P)+\mu)\gamma}{h^2}\right). \tag{4.41}$$

Combining, we get that

$$\begin{aligned}\int_0^\infty f_G(g)\sqrt{g}\int_{-\infty}^{c(g)} z f_{Z|G}(z \mid g)\,\mathrm{d}z\mathrm{d}g &= -\frac{1}{\sqrt{2\pi}}\int_0^\infty f_G(g)\sqrt{g}\exp\left(-\frac{c^2(g)}{2}\right)\mathrm{d}g \\ &= -\frac{C}{\sqrt{2\pi}}\int_0^\infty f_{\widetilde{G}}(g)\mathrm{d}g = -\frac{C}{\sqrt{2\pi}},\end{aligned} \tag{4.42}$$

where $f_{\widetilde{G}}$ is the density of GIG random variable with parameters $\widetilde{\lambda}$, $\widetilde{\chi}$ and $\widetilde{\psi}$ defined by

$$\widetilde{\lambda} = \lambda + \frac{1}{2}, \qquad \widetilde{\chi} = \chi + \frac{(\mathrm{VaR}_\alpha(P)+\mu)^2}{h^2} \quad \text{and} \quad \widetilde{\psi} = \psi + \frac{\gamma^2}{h^2}$$

and constant C is given by

$$C = \frac{\chi^{-\lambda}\left(\sqrt{\chi\psi}\right)^\lambda}{2K_\lambda\left(\sqrt{\chi\psi}\right)}\left(\frac{\widetilde{\chi}^{-\widetilde{\lambda}}\left(\sqrt{\widetilde{\chi}\widetilde{\psi}}\right)^{\widetilde{\lambda}}}{2K_{\widetilde{\lambda}}\left(\sqrt{\widetilde{\chi}\widetilde{\psi}}\right)}\right)^{-1}\exp\left(-\frac{(\mathrm{VaR}_\alpha(P)+\mu)\gamma}{h^2}\right). \tag{4.43}$$

Therefore, (4.39) reduces to

$$\mathrm{ES}_\alpha(P) = -\mu - \frac{\gamma}{\alpha}\mathbb{E}[G]\,F_{P^*}(-\mathrm{VaR}_\alpha(P)) + \frac{h}{\alpha}\frac{C}{\sqrt{2\pi}}, \tag{4.44}$$

where F_{P^*} is the cdf of $P^* \sim \mathrm{GHyp}(\mu,\gamma,h,\lambda+1,\chi,\psi)$; the constant C is given in (4.43).

4.6.2 DJIA-30 Components

Dow Jones Industrial Average components (as present on June 8, 2009)

No.	Abbrev	Company Name	No.	Abbrev	Company Names
1	MMM	3M Co	16	HD	Home Depot Inc
2	AA	Alcoa Inc	17	IBM	Intl. Business Machines Corp
3	AXP	American Express Co	18	INTC	Intel Corp
4	T	AT&T	19	JNJ	Johnson & Johnson
5	BAC	Bank of America Corp	20	JPM	JPMorgan Chase & Co
6	BA	Boeing Co	21	KFT	Kraft Foods Inc
7	CAT	Caterpillar Inc	22	MCD	McDonald's Corp
8	CVX	Chevron Corp	23	MRK	Merck & Co Inc
9	CSCO	Cisco Systems Inc	24	MSFT	Microsoft Corp
10	KO	Coca-Cola Co	25	PFE	Pfizer Inc
11	DD	El Du Pont de Nemours & Co	26	PG	Procter & Gamble Co
12	XOM	Exxon Mobil Corp	27	UTX	United Technologies Corp
13	GE	General Electric Co	28	VZ	Verizon communications Inc
14	TRV	The Travelers Companies Inc	29	WMT	Wal-Mart Stores Inc
15	HPQ	Hewlett-Packard Co	30	DIS	Walt Disney Co

5

MARC-MARS

MARC-MARS: Modeling Asset Returns via Conditional Multivariate Asymmetric Regime-Switching

Paweł Polak[a,b]

[a] *Department of Banking and Finance, University of Zurich, Switzerland*
[b] *Swiss Finance Institute*

Abstract

A regime-switching model for a multivariate set of asset returns is proposed. The model can be seen as an extension of the CCC-GARCH regime switching model of Pelletier (2006), such that the normality assumption is replaced by a (possibly special case of a) multivariate generalized hyperbolic distribution (MGHyp). In doing so, the model, like that of Pelletier (2006), accommodates the stylized facts of volatility clustering and non-constant correlations between asset returns. However, the new model also allows for asymmetry (possibly differing in each asset), and excess kurtosis in the returns—these being the most prominent and ubiquitous stylized facts of asset returns, and the most crucial for realistic risk management purposes. The multivariate predictive distribution is a discrete mixture of MGHyp distributions, and so weighted sums of marginals are themselves discrete mixture of GHyp, and thus tractable, enabling, e.g., portfolio optimization. To accomplish joint likelihood estimation of all the model parameters, a new, two-stage, EM algorithm is developed for estimation. This is coupled with shrinkage via a quasi-Bayesian prior, which enhances both estimation ease, and forecast quality. The new model is demonstrated to outperform all special cases in terms of in sample fit and out-of-sample forecasts.

Keywords: CCC; Density Forecasting; EM-Algorithm; Fat Tails; GARCH; Laplace Distribution; Markov-Switching; Multivariate Generalized Hyperbolic Distribution.

JEL Classification: C51; C53; C58; G11; G17.

5.1 Introduction

We consider modeling asset returns via a conditional multivariate asymmetric regime-switching model, or, in short, MARC-MARS. The approach originates from the Constant Conditional Correlation (CCC) model of Bollerslev (1990) and its extension to the Regime Switching Dynamic Correlation (RSDC) model of Pelletier (2006). The latter structure is appealing because it allows switching among several correlation matrices to accommodate the stylized fact that correlations are not constant through time, such as during periods of severe market downturns, in which correlations are larger than during periods of relative tranquility. The RSDC model was demonstrated by Paolella (2013) to indeed lead to improved density forecasts over

the CCC model, but is in turn outperformed by simpler non-Gaussian multivariate models proposed therein. In this paper, we show how the RSDC model can be extended to the non-Gaussian case; in particular to support use of the multivariate generalized hyperbolic (MGHyp) distribution. In doing so, the MARC-MARS structure nests the RSDC model, as well as the iid MGHyp model (McNeil et al., 2005, Ch. 3), and univariate GARCH in which the innovations sequence is taken to be GHyp or one of its special or limiting cases (e.g., normal, Student's t, NIG, Hyperbolic Skew Student's t, etc); see, among others, Jensen and Lunde (2001) and Aas and Haff (2006a).

The proposed model also generalizes the class of common market factor non-Gaussian returns (COMFORT) models introduced in Paolella and Polak (2013b), hereafter PP13. An interesting feature of the latter is the introduction of a stochastic term which can be interpreted as a market factor, common to each of the assets. This term is responsible for modeling news which affects the distribution of current period returns. While the COMFORT is a very flexible model (which outperforms all its special cases previously used), its dynamics cannot handle pronounced changes in the correlation structure related to switching, e.g., between bull and bear periods in the market, because the dependency matrix is specified as CCC. In this paper, we relax this assumption and allow for regime switches in the dependency matrix.

Crucially, the most general model proposed is straightforward to estimate, even with a large number of assets, owing to a new, two-step method for likelihood maximization. As detailed below in the empirical section, the switching model, with two regimes, significantly outperforms the single-regime case (while the use of three regimes is inferior to the use of two). To further motivate the use of our proposed new, likelihood-based estimation algorithm, issues related to its convergence properties are studied.

An interesting feature of the MARC-MARS model is that it allows incorporation of three types of asymmetries, and distinguishes between their effects: (i) skewness in the conditional density which drives the individual asset news effects; (ii) asymmetry in individual asset volatility dynamics; and, (iii) asymmetry in the dynamics of the dependency structure which is responsible for different persistence of the state-specific dependency between assets. The empirical exercise details the importance of each of these effects. Our findings confirm that each one of them has its own positive effect on the quality of the density forecasts.

The remainder of the paper is as follows. Section 5.2 states the model, while Section 5.3 discusses the proposed method of estimation and some of it properties. Section 5.4 details an empirical study, and Section 5.5 provides some concluding remarks. An Appendix gathers the technical results related to the new EM algorithm.

5.2 Model

Let $\mathbf{Y}_t = (Y_{t,1}, Y_{t,2}, \ldots, Y_{t,K})'$ denote a return vector of K financial assets at time t, for $t = 1, 2, \ldots, T$. The equally spaced (ignoring the weekend effect for daily data) realization of the return vector is denoted by $\mathbf{Y} = [\mathbf{Y}_1 \mid \mathbf{Y}_2 \mid \cdots \mid \mathbf{Y}_T]$ and the information set at time t, defined as the history of returns, is denoted by $\mathbf{\Phi}_t = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_t\}$. In what follows we assume that $\mathbf{Y}_t$

has a COMFORT representation, from PP13, given by

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu} + \boldsymbol{\gamma} G_t + \boldsymbol{\varepsilon}_t, \quad \text{with} \\ \boldsymbol{\varepsilon}_t &= \mathbf{H}_t^{1/2} \sqrt{G_t} \mathbf{Z}_t, \end{aligned} \tag{5.1}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$ are column vectors in $\mathbb{R}^K$; $\mathbf{H}_t$ is a positive definite, symmetric, dispersion matrix of order K; $\mathbf{Z}_t \overset{\text{iid}}{\sim} \mathrm{N}(\mathbf{0}, \mathbf{I}_K)$ is a sequence of independent and identically distributed (iid) normal random variables and $G_t \sim \mathrm{GIG}(\lambda, \chi, \psi)$ are iid mixing random variables, $t = 1, 2, \ldots, T$, independent of $\mathbf{Z}_t$, with typical GIG (generalized inverse Gaussian) density given by

$$f_{G|\mathbf{Y}}(x; \lambda, \chi, \psi) = \frac{\chi^{-\lambda} \left(\sqrt{\chi\psi}\right)^{\lambda}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} x^{\lambda-1} \exp\left(-\frac{1}{2}\left(\chi x^{-1} + \psi x\right)\right), \quad x > 0; \tag{5.2}$$

$K_\lambda(x)$ is the modified Bessel function of the third kind (and not to be confused with K, the number of assets), given by

$$K_\lambda(x) = \frac{1}{2} \int_0^\infty t^{\lambda-1} \exp\left(-\frac{x}{2}\left(t + t^{-1}\right)\right) \mathrm{d}t, \quad x > 0. \tag{5.3}$$

and $\chi > 0$, $\psi \geq 0$ if $\lambda < 0$; $\chi > 0$, $\psi > 0$ if $\lambda = 0$; and $\chi \geq 0$, $\psi > 0$ if $\lambda > 0$. The same MGHyp distribution arises from the parameter constellation $(\lambda, \chi/c, c\psi, \boldsymbol{\mu}, c\mathbf{H}_t, c\boldsymbol{\gamma})$ for any $c > 0$.

The conditional, positive definite, dispersion matrix $\mathbf{H}_t$ is decomposed as

$$\mathbf{H}_t \equiv \mathbf{S}_t \boldsymbol{\Gamma}_t \mathbf{S}_t, \tag{5.4}$$

where $\mathbf{S}_t$ is a diagonal matrix composed of the strictly positive conditional scale terms $s_{k,t}$, $k = 1, \ldots, K$, and $\boldsymbol{\Gamma}_t$ is a dependency matrix. The univariate scale terms $s_{k,t}$ are modeled by a GARCH-type process. The obvious starting point is the GARCH(1,1) model

$$s_{k,t}^2 = \omega_k + \alpha_k \varepsilon_{k,t-1}^2 + \beta_k s_{k,t-1}^2, \tag{5.5}$$

where $\varepsilon_{k,t} = y_{k,t} - \mu_k - \gamma_k G_t$ is the kth element of the $\boldsymbol{\varepsilon}_t$ vector in (5.1), and $\omega_k > 0$, $\alpha_k \geq 0$, $\beta_k \geq 0$, for $k = 1, 2, \ldots, K$. However, models which can capture a leverage effect typically yield better forecasts (see, e.g., Mittnik and Paolella (2000) and the references therein). In particular, the A-PARCH(1,1) model of Ding et al. (1993), given by

$$s_{k,t}^{\delta_k} = \omega_k + \alpha_k \left(|\varepsilon_{k,t-1}| - \eta_k \varepsilon_{k,t-1}\right)^{\delta_k} + \beta_k s_{k,t-1}^{\delta_k}, \quad |\eta_k| < 1, \quad \delta_k > 0, \tag{5.6}$$

has been shown to be a more effective filter than (5.5) because (i) parameter δ is typically estimated to be closer to 1 than 2 (see the detailed empirical results in Broda et al., 2012, and the references therein), and (ii) parameter η serves to capture asymmetry in the scale-term response to the last period innovation; see, e.g., Mittnik and Paolella (2000) and the references therein. In the empirical section below, we confirm that use of (5.6) in conjunction with the MARC-MARS model yields forecasts which are overall superior to use of (5.5).

For the dependency matrix, the CCC model entails using the time invariant matrix $\mathbf{\Gamma}_t = \mathbf{\Gamma}$ leading to a conditional MGhyp distribution of the returns. Dynamics in the dependency matrix are obtained by the RSDC model with N regimes, whereby

$$\mathbf{\Gamma}_t = \sum_{n=1}^{N} \mathbf{1}_{\{\Delta_t = n\}} \mathbf{\Gamma}_n, \tag{5.7}$$

where $\mathbf{1}_{\{\}}$ is an indicator function; Δ_t is a latent random variable governed by a first order Markov chain, independent of $\mathbf{Z}_t$ and G_t, which can take one of N possible values; and $\mathbf{\Gamma}_n$, $n = 1, \ldots, N$, are state specific $K \times K$ positive definite correlation matrices such that $\mathbf{\Gamma}_n \neq \mathbf{\Gamma}_m$ for $n \neq m$. The probability law governing Δ_t is defined by its time-invariant transition probability matrix, denoted by $\mathbf{\Pi} = [\pi_{n,m}]_{n,m=1,\ldots,N}$, where $\pi_{n,m}$ is the probability of going from state n in period t to state m in period $t+1$. Hence in the RSDC version of the model the conditional distribution of the returns is a discrete mixture of MGHyp.

For notational convenience later, we collect the parameters of the model into three vectors (process, distribution, and correlation)

$$\begin{aligned} \boldsymbol{\theta}_P &= \left(\boldsymbol{\mu}', \boldsymbol{\gamma}', \boldsymbol{\omega}', \boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\eta}'\right)', \quad \boldsymbol{\theta}_D = (\lambda, \chi, \psi)', \text{ and} \\ \boldsymbol{\theta}_C &= \left(\operatorname{vech}\left(\mathbf{\Gamma}_1\right)', \ldots, \operatorname{vech}\left(\mathbf{\Gamma}_N\right)', \operatorname{vec}_{N-1}\left(\mathbf{\Pi}\right)'\right)', \end{aligned} \tag{5.8}$$

where $\boldsymbol{\omega}, \boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\eta}$ are K-dimensional vectors of parameters from (5.6) (we set the power parameters δ_k, for $k = 1, \ldots, K$, to one a prior to estimation, and in case of GARCH$(1,1)$ dynamics, in (5.5), the asymmetry parameters in $\boldsymbol{\eta}$ reduce to zero); λ, χ and ψ are the GIG parameters; $\operatorname{vech}\left(\mathbf{\Gamma}_n\right)$, for $n = 1, \ldots, N$, denotes a column vector of the elements above the main diagonal of matrix $\mathbf{\Gamma}_n$; and $\operatorname{vec}_{N-1}\left(\mathbf{\Pi}\right)$ is a column vector of first $N-1$ columns of the transition probabilities matrix $\mathbf{\Pi}$ (but for just the CCC model, $\boldsymbol{\theta}_C$ reduces to $\operatorname{vech}\left(\mathbf{\Gamma}\right)$).

In the empirical application we restrict our attention to the multivariate asymmetric Laplace distribution (MALap), with density $f_{\mathbf{Y}_t}(\mathbf{y}; \lambda, \boldsymbol{\mu}, \mathbf{H}_t, \boldsymbol{\gamma})$ given by

$$\frac{2\exp\left\{(\mathbf{y}-\boldsymbol{\mu})'\,\mathbf{H}_t^{-1}\boldsymbol{\gamma}\right\}}{(2\pi)^{K/2}\Gamma\left(\lambda\right)|\mathbf{H}_t|^{1/2}} \left(\frac{m_t}{2+\boldsymbol{\gamma}'\mathbf{H}_t^{-1}\boldsymbol{\gamma}}\right)^{\lambda/2-K/4} K_{\lambda-K/2}\left(\sqrt{m_t\left(2+\boldsymbol{\gamma}'\mathbf{H}_t^{-1}\boldsymbol{\gamma}\right)}\right), \tag{5.9}$$

where $\mathbf{H}_t$ is given by (5.4) and $m_t = (\mathbf{y}-\boldsymbol{\mu})'\,\mathbf{H}_t^{-1}\,(\mathbf{y}-\boldsymbol{\mu})$. This is a special case of the MGHyp distribution in which G_t is iid gamma distributed with shape parameter $\lambda > 0$ and unit scale parameter. PP13 argue that the MGHyp is too flexible for returns data and exhibits a relatively flat likelihood in some of the parameters, while the MALap does not share this problem. Moreover, it is faster and numerically more reliable to estimate, but still retains the flexibility required for modeling asset returns. Crucially, it allows for individual asset asymmetry parameters and also higher kurtosis than the normal distribution.

5.3 Two-Step Estimation for the Regime Switching Correlation Model

PP13 develop a new ECME algorithm for estimating the model under the assumption of a constant conditional correlation matrix, i.e., the single-component case. To generalize this to the N-component case, in principle, one could extend that algorithm into a nested double EM algorithm with the inner loop estimating the regime switching correlations. However, the rapid increase of the number of parameters combined with the dynamic structure in the conditional correlation matrix of the RSDC model makes even this, double EM, estimation method impractical for large K. To overcome this, a new, two-stage estimation procedure is proposed. It replaces a double iterative procedure by two separate procedures, whereby the second procedure is conducted conditional on the results from the first one.

The estimation algorithm is based on the decomposition of the complete conditional log-likelihood function into a sum of two terms: the normal log-likelihood (i.e. conditionally on the realization of the mixing random variable) and the log-likelihood function of the mixing random variable, i.e.,

$$\log L_{\mathbf{Y},\mathbf{G}}\left(\boldsymbol{\theta}_P, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C\right) = \log L_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C\right) + \log L_{\mathbf{G}}\left(\boldsymbol{\theta}_D\right), \tag{5.10}$$

Next, the normal log-likelihood can be split as in Bollerslev (1990) and Pelletier (2006), i.e.,

$$\log L_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C\right) = \log L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P\right) + \log L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C\right), \tag{5.11}$$

where $L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P\right)$ is the mean-volatility term given by

$$\begin{aligned}\log L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P\right) = -\frac{1}{2}\sum_{t=1}^{T}\Big[&K\log\left(2\pi\right) + \log\left|\mathbf{S}_t\right|^2 \\ &+ g_t^{-1}\left(\mathbf{y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma} g_t\right)'\mathbf{S}_t^{-1}\mathbf{S}_t^{-1}\left(\mathbf{y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma} g_t\right) + \log g_t\Big],\end{aligned} \tag{5.12}$$

and $L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C\right)$ is the correlation term given by

$$\log L^{\mathrm{Corr}}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C\right) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log\left|\boldsymbol{\Gamma}\right| + \boldsymbol{e}_t'\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_t - \boldsymbol{e}_t'\boldsymbol{e}_t\right], \tag{5.13}$$

where $\boldsymbol{e}_t = g_t^{-1/2}\mathbf{S}_t^{-1}\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \boldsymbol{\mu} - \boldsymbol{\gamma} g_t$ from (5.1).

Owing to the mixture structure of the MGHyp, $L_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C\right)$ is a multivariate Gaussian likelihood with a GARCH structure for the scales and a given conditional correlation model. As such, maximization of $L_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P, \boldsymbol{\theta}_C\right)$ can be done in two steps. First, with the correlation structure ignored, the GARCH parameters in (5.5) are estimated for each of the K assets separately (or concurrently with parallel computing) by maximizing $L_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P \mid \boldsymbol{\theta}_C = \mathbf{I}_K\right)$, and, in the second step, the correlation parameters $\boldsymbol{\theta}_C$ are estimated from the first-step standardized residuals. In case of constant dependency matrix, as in PP13, the $\boldsymbol{\Gamma}$ matrix is estimated by the usual empirical correlation estimator (the MLE under normality) of the standardized

residuals $g_t^{-1/2}\mathbf{S}_t^{-1}\widehat{\boldsymbol{\varepsilon}}_t$ (with the unobserved realizations of $g_t^{-1/2}$ replaced by their conditional expectations from the E-step below), this is instantaneous and trivial. When $\boldsymbol{\Gamma}_t$ has a regime switching structure the correlation step would be considerably slower. We propose to omit the correlation step and turn to the maximization of the second term in the decomposition (5.10). Given the $\boldsymbol{\theta}_P$ estimates, denoted as usual by $\widehat{\boldsymbol{\theta}}_P$, estimate mixing process parameters $\boldsymbol{\theta}_D$ by maximizing $L_{\mathbf{Y}}\big(\boldsymbol{\theta}_D \mid \widehat{\boldsymbol{\theta}}_P, \boldsymbol{\theta}_C = \mathbf{I}_K\big)$. Given all these estimates, we proceed with the next E-step update of the unobserved mixing random variable $\mathbf{G}$ (still with the ignored correlation structure) and continue to iterate until convergence. The estimates of the correlation dynamics, $\boldsymbol{\theta}_C$, are obtained in the second stage by a separate EM algorithm.

Stage-I E-step: Calculate $\mathbb{E}\big[\log L_{\mathbf{Y},\mathbf{G}} \mid \mathbf{Y}; \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_D, \boldsymbol{\theta}_C = \mathbf{I}_K\big]$.

The log-likelihood function (5.12) is linear with respect to g_t and g_t^{-1}. Hence, the E-step involves replacing unobserved realizations of G_t and G_t^{-1} in (5.10) by their conditional expectations $\mathbb{E}\left[G_t^{\pm 1} \mid \mathbf{Y}; \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_D, \boldsymbol{\theta}_C = \mathbf{I}_K\right]$.

Stage-I CM1-step: Update $\boldsymbol{\theta}_P$ by computing

$$\arg\max_{\boldsymbol{\theta}_P} \log L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P\right), \tag{5.14}$$

where $L^{\mathrm{MV}}_{\mathbf{Y}|\mathbf{G}}\left(\boldsymbol{\theta}_P\right)$ is a Gaussian likelihood with zero correlation, so we can estimate the parameters of each asset, $(\mu_k, \gamma_k, \omega_k, \alpha_k, \beta_k)$, separately by maximizing the corresponding likelihood function.

Stage-I CM2-step: Given the CM1-step estimates of $\boldsymbol{\theta}_P$, obtain new estimates of $\boldsymbol{\theta}_D$ by maximizing the incomplete data log-likelihood function, i.e., compute

$$\arg\max_{\boldsymbol{\theta}_D} \log L_{\mathbf{Y}}\left(\boldsymbol{\theta}_D \mid \widehat{\boldsymbol{\theta}}_P, \boldsymbol{\theta}_C = \mathbf{I}_K\right). \tag{5.15}$$

Iterate the above steps until convergence.

Stage-II Conditional on $\widehat{\boldsymbol{\theta}}_P$ and $\widehat{\boldsymbol{\theta}}_D$, estimate the remaining parameters as

$$\arg\max_{\boldsymbol{\theta}_C} \log L_{\mathbf{Y}}\big(\boldsymbol{\theta}_C \mid \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_D\big). \tag{5.16}$$

The Stage-II step replaces the correlation step, from the original algorithm in PP13, and is itself an EM-algorithm, conducted conditionally on the other parameters estimates. The algorithm generalizes the (Hamilton, 1993, 1994) method for time series with changes in regime under the normality assumption to the MGHyp distribution case. It is embedded with the smoothed inference algorithm for the regime probabilities from Kim (1994). In order to reduce the estimation error and to increase the forecasting performance of the model, we augment the estimators with a quasi-Bayesian prior along the lines of Hamilton (1991) and Paolella (2013).

The Stage-II EM algorithm then consists of the following 5 steps:

(II-1) Set iteration count $\ell = 1$ and choose the starting values for $\boldsymbol{\theta}_C^{[1]} = \left[\boldsymbol{\Gamma}_1^{[1]}, \ldots, \boldsymbol{\Gamma}_N^{[1]}, \boldsymbol{\Pi}^{[1]}\right]$.

(II-2) Recursively calculate the probabilities $\boldsymbol{\xi}_{t+1|t}^{[\ell]}$, $\boldsymbol{\xi}_{t|t}^{[\ell]}$ and a smoothed version $\boldsymbol{\xi}_{t|T}^{[\ell]}$, using

$$\boldsymbol{\eta}_t^{[\ell]} = \begin{bmatrix} f\left(\mathbf{y}_t \mid \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{t-1}, \Delta_t = 1; \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_D, \boldsymbol{\Gamma}_1^{[\ell]}\right) \\ f\left(\mathbf{y}_t \mid \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{t-1}, \Delta_t = 2; \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_D, \boldsymbol{\Gamma}_2^{[\ell]}\right) \\ \vdots \\ f\left(\mathbf{y}_t \mid \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{t-1}, \Delta_t = N; \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_D, \boldsymbol{\Gamma}_N^{[\ell]}\right) \end{bmatrix},$$

and

$$\boldsymbol{\xi}_{t|t}^{[\ell]} = \frac{\boldsymbol{\xi}_{t|t-1}^{[\ell]} \odot \boldsymbol{\eta}_t^{[\ell]}}{\mathbf{1}_N' \left(\boldsymbol{\xi}_{t|t-1}^{[\ell]} \odot \boldsymbol{\eta}_t^{[\ell]}\right)},\ \boldsymbol{\xi}_{t+1|t}^{[\ell]} = \boldsymbol{\Pi}^{[\ell]} \boldsymbol{\xi}_{t|t}^{[\ell]},\ \boldsymbol{\xi}_{t|T}^{[\ell]} = \boldsymbol{\xi}_{t|t}^{[\ell]} \odot \left\{\boldsymbol{\Pi}' \left[\boldsymbol{\xi}_{t+1|T}^{[\ell]} \div \boldsymbol{\xi}_{t+1|t}^{[\ell]}\right]\right\}, \quad (5.17)$$

where $\boldsymbol{\xi}_{s|t}^{[\ell]}$ denotes an N-dimensional vector of probabilities of observing each of the N states at time s, based on the data obtained through date t, and based on knowledge of the population parameters (5.8); $\mathbf{1}_N$ is an $N \times 1$ vector of ones; and $\odot$ and $\div$ denote the element by element product and division, respectively.

(II-3) Update the estimates of the correlation matrices by

$$\tilde{\boldsymbol{\Gamma}}_n^{[\ell+1]} = \frac{a_n \mathbf{B}_n + \sum_{t=1}^T g_t^{-1} \mathbf{S}_{n,t}^{-1} \left(\mathbf{y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\gamma}} g_t\right) \left(\mathbf{y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\gamma}} g_t\right)' \mathbf{S}_{n,t}^{-1} \xi_{n,t|T}^{[\ell]}}{a_n + \sum_{t=1}^T \xi_{n,t|T}^{[\ell]}}, \quad (5.18)$$

(the tilde indicating that the matrix is not yet standardized to be an actual correlation matrix; see (5.21) below), where fixed quantities $a_n \geq 0$; and $\mathbf{B}_n$ positive definite $K \times K$ matrices with ones on the diagonal and off-diagonal elements between -1 and 1, represent the weight and the values of the nth prior correlation matrix, respectively, $n = 1, \ldots, N$, as discussed below in Section 5.4; $\mathbf{S}_{n,t}$ are $K \times K$ diagonal matrices with scale terms generated by equation (5.5), with first-step estimates of $\boldsymbol{\theta}_P$ and regime-specific estimates of G_t, denoted g_t, from the previous iteration ℓ; $\xi_{n,t|T}^{[\ell]}$ denotes the nth element of the probability vector $\boldsymbol{\xi}_{t|T}^{[\ell]}$ from (5.17); and all hatted parameters come from the first estimation stage.

As the latent values g_t and g_t^{-1} enter linearly in (5.18), their estimates are given by the conditional expectation of G_t and G_t^{-1}, respectively, given (i) the observed data $\mathbf{Y}_t$ and the $\mathbf{S}_{n,t}$ matrix; (ii) the parameter estimates $\widehat{\boldsymbol{\theta}}_P$ and $\widehat{\boldsymbol{\theta}}_D$; (iii) the previous iteration value, $\boldsymbol{\theta}_C^{[\ell]}$; and (iv) expectations of G_{t-1} (see Remark (ii) below). Paralleling Stage-I E-step, we get

$$\left(G_t \mid \mathbf{Y}_t, \mathbf{S}_{n,t}, G_{t-1}, \Delta_t = n; \widehat{\boldsymbol{\theta}}_P, \widehat{\boldsymbol{\theta}}_D, \boldsymbol{\theta}_C^{[\ell]}\right) \sim \text{GIG}\left(\lambda^{[\ell]}, \chi^{[\ell]}, \psi^{[\ell]}\right), \quad (5.19)$$

where $\lambda^{[\ell]} = \widehat{\lambda} - K/2$, $\chi^{[\ell]} = m_{n,t}^{[\ell]} + \widehat{\chi}$ and $\psi^{[\ell]} = \widehat{\psi} + \widehat{\boldsymbol{\gamma}}' \left(\mathbf{S}_{n,t} \boldsymbol{\Gamma}_n^{[\ell]} \mathbf{S}_{n,t}\right)^{-1} \widehat{\boldsymbol{\gamma}}$; and

$$m_{n,t}^{[\ell]} = \left(\mathbf{y}_t - \widehat{\boldsymbol{\mu}}\right)' \left(\mathbf{S}_{n,t} \boldsymbol{\Gamma}_n^{[\ell]} \mathbf{S}_{n,t}\right)^{-1} \left(\mathbf{y}_t - \widehat{\boldsymbol{\mu}}\right), \quad (5.20)$$

and the conditional expectations of G_t and G_t^{-1} are given by the explicit formulae for the moments of the GIG random variable.

(II-4) In order to get a correlation matrix, (5.18) needs to be rescaled. The $(\ell+1)$st iteration update of the correlation matrix cannot be used directly because the elements on the diagonal of $\mathbf{\Gamma}_n^{[\ell+1]}$ are not imposed to be unity. Instead, and as done with the RSDC model of Pelletier (2006), we set, for $n = 1, \ldots, N$,

$$\mathbf{\Gamma}_n^{[\ell+1]} = \mathbf{D}_n^{-1[\ell+1]} \tilde{\mathbf{\Gamma}}_n^{[\ell+1]} \mathbf{D}_n^{-1[\ell+1]}, \qquad \mathbf{D}_n^{[\ell+1]} = \operatorname{diag}\left(\sqrt{\tilde{\mathbf{\Gamma}}_{1,1,n}^{[\ell+1]}}, \ldots, \sqrt{\tilde{\mathbf{\Gamma}}_{K,K,n}^{[\ell+1]}} \right). \quad (5.21)$$

(II-5) Finally, each element of the transition matrix, $\mathbf{\Pi} = [\pi_{n,m}]_{n,m=1,\ldots,N}$, is updated by

$$\pi_{n,m}^{[\ell+1]} = \sum_{t=2}^{T} \xi_{m,t|T}^{[\ell]} \frac{\pi_{n,m}^{[\ell]} \xi_{n,t-1|t-1}^{[\ell]}}{\xi_{m,t|t-1}^{[\ell]}} \Bigg/ \sum_{t=2}^{T} \xi_{n,t-1|T}^{[\ell]}.$$

This sequence of steps is iterated until convergence.

Remarks:

(i) In step II-2 above, we use Hamilton's (1989) recursive filter and the smooth probabilities algorithm given in Kim (1994). What remains in this recursion is to choose starting values $\widehat{\boldsymbol{\xi}}_{1|0}$. We take the approach proposed by Hamilton (1989) and add this $K \times 1$ vector to the parameter space and estimate these initial probabilities by maximum likelihood along with the parameters in $\boldsymbol{\theta}_C$. Note that their estimation is embedded in step (2) of the EM algorithm and thus does not incur any substantial extra cost.

(ii) If the asymmetry component $\boldsymbol{\gamma}$ is used, then the scale terms $s_{k,t}$, $k = 1, \ldots, K$, for a given t, are functions of the realization of $\{G_1, \ldots, G_{t-1}\}$ through the recursive GARCH equation (5.5). Thus, in the computation of the diagonal matrix $\mathbf{S}_{n,t}^{-1}$ in step (3), we use

$$\mathbb{E}\left[G_{t-1} \mid \mathbf{Y}_{t-1}, \Delta_t = n, \mathbf{\Pi}^{[\ell]} \right] = \sum_{m=1}^{N} \pi_{n,m} \mathbb{E}\left[G_{t-1} \mid \mathbf{Y}_{t-1}, \Delta_{t-1} = m, \mathbf{\Pi}^{[\ell]} \right],$$

where $\mathbf{\Pi}^{[\ell]} = \left[\pi_{n,m}^{[\ell]} \right]_{n,m=1,\ldots,N}$ and the expectations under the sum are given by explicit formulae for moments of the GIG random variable with the parameters as in (5.19). This is the reason that the matrix $\mathbf{S}_{n,t}^{-1}$ is indexed by n, but is not a function of G_t.

(iii) Ignoring the correlation structure does not change the updates of the θ_P in the ECME algorithm. Combining this finding with some mild relative entropy condition the monotonicity of the ECME algorithm, performed on the subspace of the parameter space with the zero correlations, is preserved. Therefore, by the self consistency of the ECME algorithm, our approach maintains the consistency and asymptotic normality of estimates in the same sense as the existing two-step procedures under the normality assumption. The theorems and proofs formulating the corresponding necessary and sufficient conditions are given in Appendix 5.6.1.

5.4 Empirical Application

The data set consists of the $1,945$ daily returns of $K = 30$ components of the Dow Jones Industrial Index (DJ-30) from June 13th, 2001, to March 11th, 2009 (based on the DJ-30 composition in 2009). The ordering, abbreviations and names of the sample stocks are given in Appendix 5.6.2. Returns for each asset are computed as continuously compounded percentage returns, given by $y_{k,t} = 100 \log(p_{k,t}/p_{k,t-1})$, where $p_{k,t}$ is the price of an asset k at time t.

First, we compare the in-sample fit of the multivariate normal CCC (MN-CCC) and the MALap-CCC models and show that the latter provides a much better fit to the tails of the return distribution. Next, we discuss the MALap-CCC A-PARCH(1,1) model parameter estimates across different data windows.

Given the superior in-sample fit of our MALap-CCC model, we then compare the forecasting performance across different models. Summarizing the results, the MALap-RSDC model yields a clear prediction of which state will get realized next period and delivers the best density and covariance matrix forecasts among all considered models. In particular, we find (i) a large improvement moving from the normal to the Laplace distribution; (ii) that introducing asymmetries in the density function and the correlation dynamics in combination with shrinkage further improves the forecasting performance; (iii) that the use of the optimal amount of shrinkage increases the "separation" of the regime forecasts; (iv) that the use of the more general A-PARCH(1,1) model (5.6) instead of (5.5) further increases forecasting performance; and (v) that the MALap-RSDC model results in better predictions of the matrix of the second moments, compared to the MN-CCC and MN-RSDC models, or the simpler MLap-CCC and MALap-CCC models from PP13.

5.4.1 In-Sample Performance

The in-sample fit is evaluated based on the Q-Q-plots of the sample quantiles of the standardized residuals versus the theoretical quantiles from a normal distribution, where the standardized residuals, for the MN-CCC model and for the MALap-CCC model, are given by

$$\widehat{\mathbf{H}}_t^{-1/2}\big(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}}\big) \quad \text{and} \quad \widehat{G}_t^{-1/2}\widehat{\mathbf{H}}_t^{-1/2}\big(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\gamma}}\widehat{G}_t\big), \tag{5.22}$$

respectively. Where $\widehat{G}_t$ are the imputed values of G_t returned from the EM algorithm, $\widehat{\mathbf{H}}_t$ are fitted conditional dispersion matrices and other hatted entries denote parameter estimates. Observe that both sets of residuals in (5.22), in particular the latter, are assumed to be Gaussian under each of the assumed models. The set of all figures for all the models and for 30 assets is available upon request. They confirm our previous findings that moving from Gaussian models to any of the MALap models improves the tail fit.

Figure 5.1 shows the six components of $\widehat{\boldsymbol{\theta}}_P$ from the MALap-CCC A-PARCH(1,1) model. The parameter estimates are obtained for 946 moving windows of $1,000$ observations. The elements of the estimated location vector $\widehat{\boldsymbol{\mu}}$ in the top left panel initially are nearly all positive. Towards the end of the sample the $\widehat{\boldsymbol{\mu}}$ are more equally distributed around zero. By contrasting the estimates of $\boldsymbol{\omega}$ (top right panel), $\boldsymbol{\alpha}$ (middle left panel) and $\boldsymbol{\beta}$ (middle right panel), one

can identify periods of a low (high) constant volatility component with a high (low) persistence of the assets scale term, i.e., β_k close to 1; and a smaller (larger) impact of the last period shock on the assets scale term, i.e., α_k close to zero. These periods correspond to the periods of low (high) volatility in the data. The estimates of $\boldsymbol{\gamma}$ (bottom left panel) are predominantly negative, indicating fatter left tails of the returns distribution. The $\boldsymbol{\gamma}$ vector is responsible for the impact of the common factor G_t on each asset. Negative signs of $\boldsymbol{\gamma}$ components suggest that G_t captures mostly negative news. Finally, the estimated A-PARCH asymmetry coefficients $\boldsymbol{\eta}$ (bottom right) are positive, indicating an asymmetry in the scale response to the last period shock. This result is exactly in line with the leverage effect.

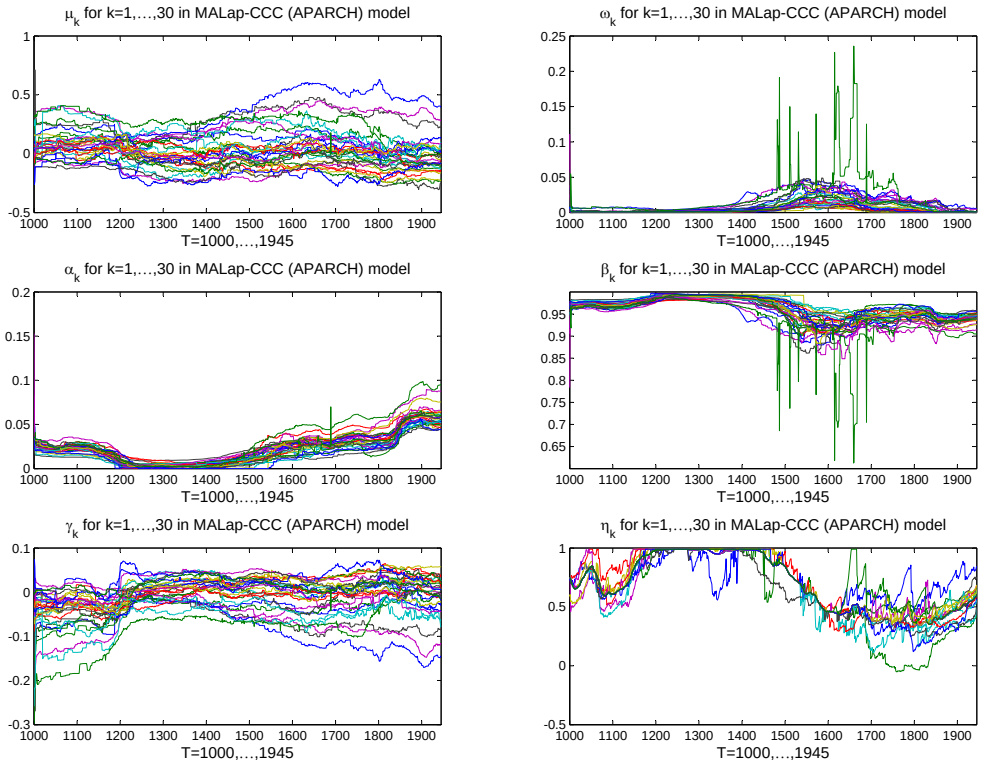

Figure 5.1: Estimates of $\boldsymbol{\theta}_P$, for the MALap-CCC A-PARCH(1,1) model; for 30 assets plotted for 946 windows of 1,000 observations. Row-wise $\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\omega}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\eta}}$.

5.4.2 Density Forecasting Performance Comparison

In this section we compare the models from the literature (the CCC model of Bollerslev (1990), the DCC model of Engle and Sheppard (2002) and the RSDC model of Pelletier (2006), all denoted with a prefix MN- for Multivariate Normal distribution of the innovations) and five multivariate Laplace based models proposed in this paper (two with symmetric Laplace distribution MLap-CCC and MLap-RSDC; and two with asymmetric Laplace distribution MALap-

CCC and MALap-RSDC). The scale term dynamics from (5.5) and (5.6) are compared for two asymmetric models MALap-CCC and MALap-RSDC, and for the MALap-RSDC A-PARCH model we investigate the optimal number of regimes. Finally, we compare the model predictions of the matrix of second moments.

Our interest centers on the quality of one-step ahead predictions of the return vector density. For this purpose, we use the normalized sum of the realized predictive log-likelihood, as proposed and discussed in Paolella (2013), given by

$$S_T(\mathcal{M}) = \frac{1}{T}\sum_{t=1}^{T} \pi_t(\mathcal{M}), \tag{5.23}$$

where

$$\pi_t(\mathcal{M}) = \log f_{t+1|t}^{\mathcal{M}}(\mathbf{Y}_{t+1} \mid \boldsymbol{\theta}), \tag{5.24}$$

and $\mathcal{M}$ indicates the model.

For the RSDC models the correlation matrix is stochastic. Hence, we extend the above definition of $\pi_t(\mathcal{M})$ by replacing (5.24) with a conditional expectation analogue, i.e.,

$$\pi_t(\mathcal{M}) = \log \mathbb{E}_t\left[f_{t+1|t}^{\mathcal{M}}\left(\mathbf{Y}_{t+1} \mid \widehat{\boldsymbol{\theta}}\right)\right] = \log \sum_{n=1}^{N} \xi_{n,t+1|t}\, f_{t+1|t}^{\mathcal{M}}\left(\mathbf{y}_{t+1} \mid \Delta_{t+1} = n; \widehat{\boldsymbol{\theta}}\right), \tag{5.25}$$

where $\xi_{n,t+1|t}$ is the nth element of the vector $\boldsymbol{\xi}_{t+1|t}$, given in (5.17). As the Markov structure does not affect the parameters of the mixing random variable G_t, one can equivalently replace the correlation matrix by its conditional expectation and compute $\pi_t(\mathcal{M}, \nu)$ directly from the MALap density in (5.9) with a correlation matrix given by $\widehat{\mathbf{\Gamma}}_{t+1} = \sum_{n=1}^{N} \xi_{n,t+1|t}\mathbf{\Gamma}_n$. To show the equivalence of using $\mathbb{E}_t\left[f_{t+1|t}^{\mathcal{M}}\left(\mathbf{Y}_{t+1} \mid \widehat{\boldsymbol{\theta}}\right)\right]$ as given in (5.25) and

$$f_{t+1|t}^{\mathcal{M}}\left(\mathbf{y}_{t+1} \mid \widehat{\boldsymbol{\theta}}\right) = f_{t+1|t}^{\mathcal{M}}\left(\mathbf{y}_{t+1} \mid \widehat{\boldsymbol{\theta}}, \mathbb{E}_t\left[\mathbf{\Gamma}_{t+1}\right]\right), \tag{5.26}$$

let $G \sim \mathrm{GIG}(\lambda, \chi, \psi)$ and $(\mathbf{X} \mid G = g) \sim \mathrm{N}(\mathbf{m}, \mathbf{\Sigma}_n)$ and define $\mathbf{m} = \boldsymbol{\mu} - \boldsymbol{\gamma} g$. Then

$$\begin{aligned}\sum_{n=1}^{N} \xi_n \left[\int f_{\mathbf{X}|G=g}(y) f_G(g) \mathrm{d}g\right] &= \int \sum_{n=1}^{N} \xi_n f_{\mathbf{X}|G=g}(y; \mathbf{m}, \mathbf{\Sigma}_n) f_G(g) \mathrm{d}g \\ &= \int f_{\mathbf{X}|G=g}\left(y \mid g; \mathbf{m}, \sum_{n=1}^{N} \xi_n(\mathbf{\Sigma}_n)\right) f_G(g) \mathrm{d}g,\end{aligned}$$

where the last step follows from the basic properties of normal random variables.

Using (5.23), we analyze the forecasting performance of the aforementioned models. We estimate each model on 946 moving windows of length $\nu = 1,000$ using the DJ-30 data set.

The MALap-RSDC model is the most general model. It includes all other considered models as special or limiting cases. However, the corresponding estimates can be biased due to the misspecification error and the estimation error might be large due to the rich structure of the model. In order to find an optimal trade-off between these two errors, we introduce two types

of shrinkage in the estimation algorithm: (i) on the skewness parameter $\boldsymbol{\gamma}$, by penalizing the first step likelihood function with the penalty term $\vartheta_{\boldsymbol{\gamma}} |\boldsymbol{\gamma} - \boldsymbol{\gamma}_S|$, where $\vartheta_{\boldsymbol{\gamma}} \geq 0$ is the penalty strength scalar hyper-parameter and $\boldsymbol{\gamma}_S$ is the shrinkage target; (ii) via a quasi-Bayesian prior, discussed in Section 5.3, on the regime-specific correlation matrices.

For the shrinkage we specify the strength parameters and the shrinkage targets. Regarding the targets for the skewness parameter, we shrink towards a vector of zeros ($\boldsymbol{\gamma}_S = \mathbf{0}$) and for the correlation target we refer to our one-step estimated MALap-CCC model and use the corresponding correlation matrix estimates as the shrinkage target (one can also consider an identity matrix). This choice of shrinkage targets corresponds to the simpler ($N = 1$) and limiting case ($\boldsymbol{\gamma} \to 0$) of the MALap-RSDC model.

For the strength of the prior of the skweness parameter, $\vartheta_{\boldsymbol{\gamma}}$, we investigate empirically different levels (see top left panel in Figure 5.2). For each of the correlation matrix estimators, $\widehat{\boldsymbol{\Gamma}}_n$, in our $N = 2$ setting, we take different shrinkage strength, $a_1 = 2\vartheta_{\boldsymbol{\Gamma}}$ and $a_2 = \vartheta_{\boldsymbol{\Gamma}}/2$ with $\vartheta_{\boldsymbol{\Gamma}} > 0$, a scalar hyper-parameter. Less prior strength is assigned to the second regime correlation matrix because it corresponds to more turbulent market conditions. The sources of these turbulences may be different than the one used to determine the prior (see Paolella (2013)). The hyper-parameter, $\vartheta_{\boldsymbol{\Gamma}}$, is determined based on a simulation study by analyzing the resulting forecasting performance.

Figure 5.2 consists of a collection of normalize sums of the realized predictive log-likelihood measures for different models plotted as a function of the correlation shrinkage strength parameter $\vartheta_{\boldsymbol{\Gamma}}$. For comparison purposes all the panels exhibit the same scale.

In the top left panel $S_{1945}\,(\text{MALap-RSDC}, 1,000)$ is plotted and different curves represent different levels of the skewness shrinkage strength, $\vartheta_{\boldsymbol{\gamma}}$. A very small amount of skewness shrinkage $\vartheta_{\boldsymbol{\gamma}} = 0.001$ gives the best fit for any level of $\vartheta_{\boldsymbol{\Gamma}}$. Similarly, $\vartheta_{\boldsymbol{\Gamma}} = 500$ dominates across all levels of $\vartheta_{\boldsymbol{\gamma}}$. Hence, among the MALap-RSDC models, we choose the one with skewness shrinkage $\vartheta_{\boldsymbol{\gamma}} = 0.001$ for which the optimal correlation shrinkage strength is $\vartheta_{\boldsymbol{\Gamma}} = 500$.

The top right panel in Figure 5.2 plots our best MALap-RSDC model and the normal based models. The simplest MN-CCC model performs worst. The MN-RSDC model has only a slightly better performance. The best among all the Gaussian based models is the MN-DCC model. All Gaussian based models give comparable results and the enormous domination of the MALap-RSDC model is clearly visible.

Therefore, in the remaining two panels only Laplace based models are considered. Bottom left panel collates the MALap-RSDC for $\vartheta_{\boldsymbol{\gamma}} = 0.001$ with the symmetric version of this model, the MLap-RSDC, where $\boldsymbol{\gamma}$ is set to $\mathbf{0}$ prior to the estimation. Allowing for skewness in the distribution (controlled by shrinkage) again improves the forecasting performance of the model. Even the MALap-RSDC model with $\vartheta_{\boldsymbol{\gamma}} = 0$ is better then the symmetric MLap-RSDC model (compare the top and bottom left panels in Figure 5.2), so clearly the asymmetric version of the Laplace distribution improves the forecast.

The bottom right panel compares the MALap-RSDC (for $\vartheta_{\boldsymbol{\gamma}} = 0.001$) with all the Laplace based CCC models. MALap-CCC with $\vartheta_{\boldsymbol{\gamma}} = 0$ ranks last. Even the symmetric version (MLap-CCC) performs better, but if the shrinkage on the skewness is introduced, then the MALap-CCC

model with $\vartheta_{\boldsymbol{\gamma}} = 0.001$ catches up and, not surprisingly, forms a limiting case for MALap-RSDC model with $\vartheta_{\boldsymbol{\gamma}} = 0.001$ and $\vartheta_{\boldsymbol{\Gamma}} \to \infty$. Although $\vartheta_{\boldsymbol{\Gamma}} = 500$ for an optimal correlation shrinkage strength seems a lot, this plot shows how slowly the MALap-RSDC converges to the MALap-CCC (for $\vartheta_{\boldsymbol{\Gamma}} = 8000$ the difference is still visible). Again, starting from some level of correlations shrinkage, $\vartheta_{\boldsymbol{\Gamma}}$, the MALap-RSDC (with $\vartheta_{\boldsymbol{\gamma}} = 0.001$) starts to dominate all other models in the panel.

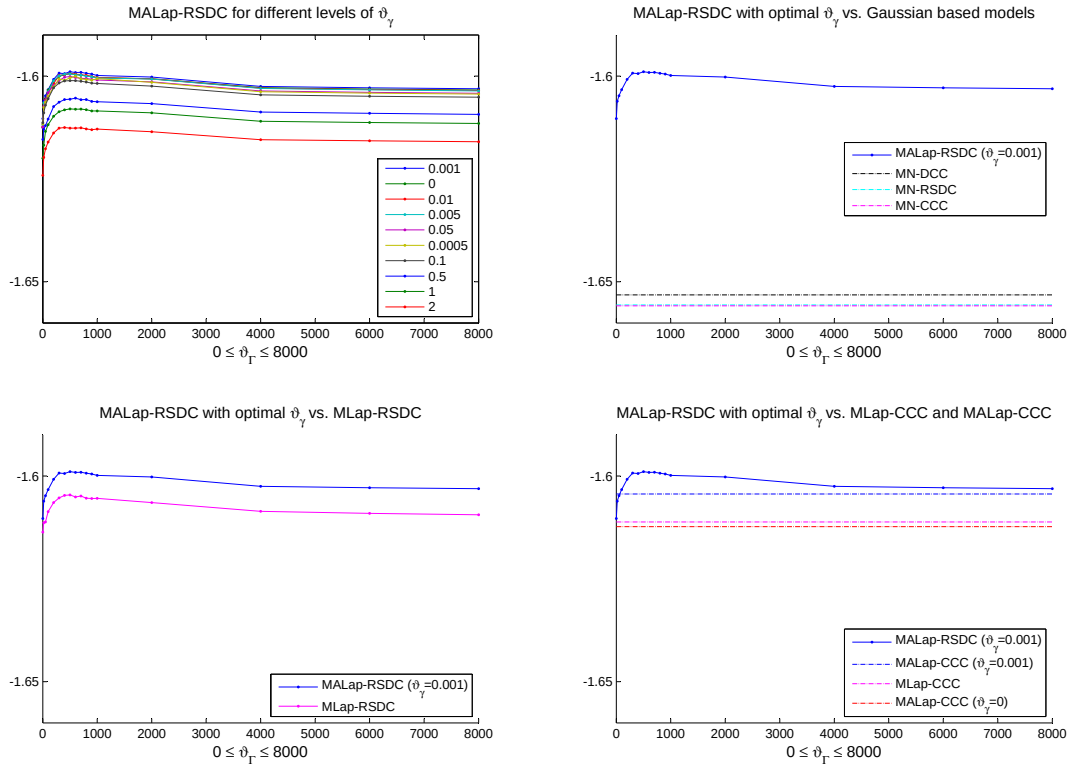

Figure 5.2: Normalized sum of the realized predictive log-likelihood measures, $S_{1945}(\mathcal{M})$, divided by number of assets, for different models, $\mathcal{M}$, plotted as a function of the correlations shrinkage strength parameter $\vartheta_{\boldsymbol{\Gamma}}$. **Top left:** $\mathcal{M}$ = MALap-RSDC; different curves represent different levels of the skewness shrinkage strength, $\vartheta_{\boldsymbol{\gamma}}$ (see the legend). **Top right:** MALap-RSDC model, for $\vartheta_{\boldsymbol{\gamma}} = 0.001$, vs. Gaussian based models (MN-CCC, MN-RSDC and MN-DCC). **Bottom left:** MALap-RSDC, for $\vartheta_{\boldsymbol{\gamma}} = 0.001$, vs. the symmetric version - the MLap-RSDC model. **Bottom right:** MALap-RSDC ($\vartheta_{\boldsymbol{\gamma}} = 0.001$) vs. the Laplace based CCC models.

Given that our MALap-RSDC with GARCH(1,1) dynamics model outperforms all of the competitors, we investigate if the introduction of more general scale term dynamics can further increase the performance of our best model. For this purpose, we replace the simple GARCH(1,1) model, from equation (5.5), by a more general A-PARCH(1,1) structure, from equation (5.6) and conduct an analogous out-of-sample exercise. Along the estimation of the A-PARCH(1,1) model we follow the standard procedure and fix the power parameter, $\delta_k = 1$. The

results are given in Figure 5.3. A-PARCH(1,1) dynamics clearly outperform the GARCH(1,1) model. Slightly more skewness shrinkage strength, $\vartheta_{\boldsymbol{\gamma}}$, is optimal and the difference in the performance between models with different $\vartheta_{\boldsymbol{\gamma}}$ has diminished in the A-PARCH(1,1) model.

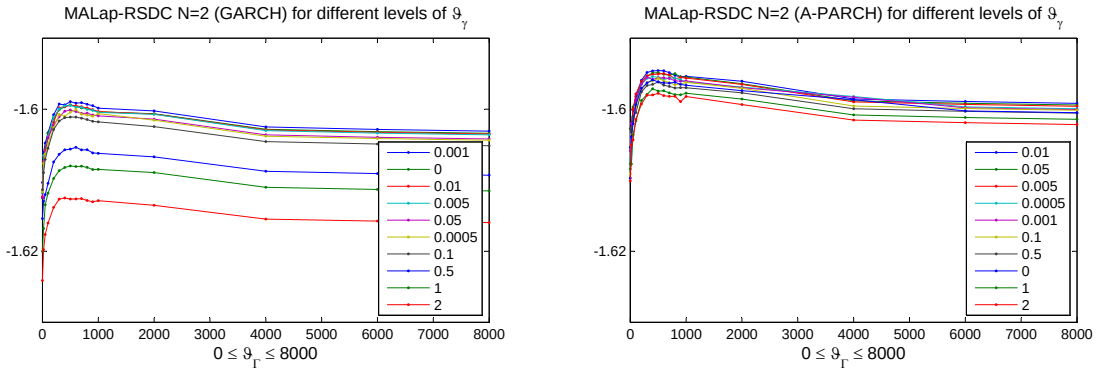

Figure 5.3: Normalized sum of the realized predictive log-likelihood measures, $S_{1945}(\mathcal{M})$, divided by number of assets, for different models, $\mathcal{M}$, plotted as a function of the correlations shrinkage strength parameter $\vartheta_{\boldsymbol{\Gamma}}$; different curves represent different levels of the skewness shrinkage strength, $\vartheta_{\boldsymbol{\gamma}}$ (see the legend). **Left:** $\mathcal{M}$ = MALap-RSDC with GARCH dynamics; **Right:** $\mathcal{M}$ = MALap-RSDC with A-PARCH(1,1) dynamics.

One might be interested in investigating if the two regimes are optimal to explain the dynamics of the correlation matrix. In order to answer this question we run our forecasting simulation exercise also for three ($N = 3$) regimes. Figure 5.4 compares the forecasting performance of the MALap-RSDC (A-PARCH) model with two ($N = 2$) and three ($N = 3$) regimes. Eventually, three regimes are able to forecast equally good (but not better) as two regimes, but the important drawback of $N = 3$ case is that to get comparable results one has to increase the shrinkage strength to $\vartheta_{\boldsymbol{\Gamma}} = 1,000$ which is twice as much as in the $N = 2$ case. Hence, one gets more regimes but these have to be forced to be more similar to each other with respect to the corresponding correlation matrices. Since $N = 3$ does not introduce any substantial increase in the forecasting performance and requires much more shrinkage, we conclude that the $N = 2$ case is sufficient and optimal as it requires much less parameters to estimate (hence with lower estimation error) and is faster in the computation.

The proposed shrinkage of the correlation matrices does not only improve the forecasting performance of the MALap-RSDC model but also has an impact on the separation of the one-step ahead regimes forecasts. Top panels of Figure 5.5 show how the out-of-sample predictions of the probabilities which state is going to realize next, $\boldsymbol{\xi}_{t+1|t}$, evolve when we incorporate shrinkage into the model. Shrinkage increases the separation between the states by shifting the probabilities closer to the boundary values (0 and 1). When we do a vertical comparison between top and bottom panels in figure 5.5 it is clear that the second regime is identified with periods of more volatile returns.

We denote by $\boldsymbol{\Sigma}_{t+d|t}$ the d-day ahead forecast of the covariance matrix, for the Laplace

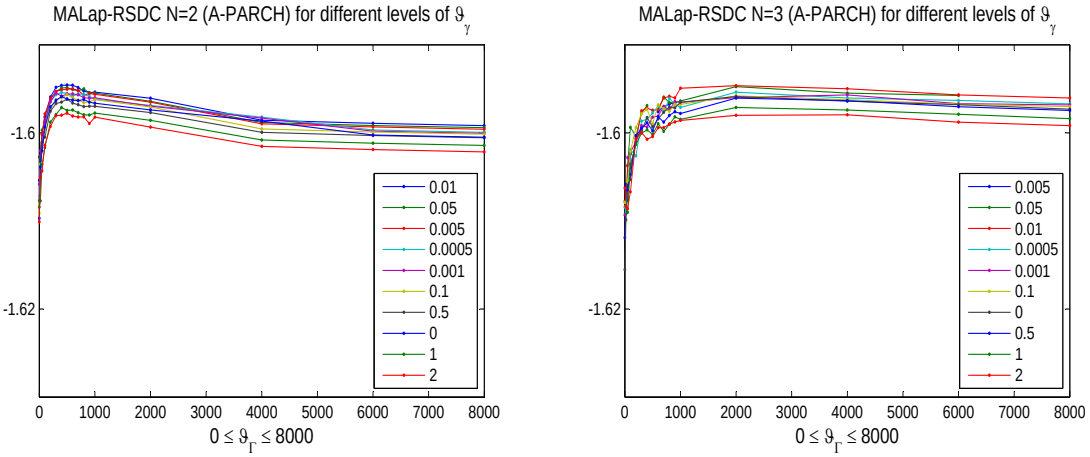

Figure 5.4: Normalized sum of the realized predictive log-likelihood measures, $S_{1945}(\mathcal{M})$, divided by number of assets, for different models, $\mathcal{M}$, plotted as a function of the correlations shrinkage strength parameter $\vartheta_{\mathbf{\Gamma}}$; different curves represent different levels of the skewness shrinkage strength, $\vartheta_{\boldsymbol{\gamma}}$ (see the legend). **Left:** $\mathcal{M}$ = MALap-RSDC with $N = 2$ and A-PARCH dynamics; **Right:** $\mathcal{M}$ = MALap-RSDC with $N = 3$ and A-PARCH(1,1) dynamics.

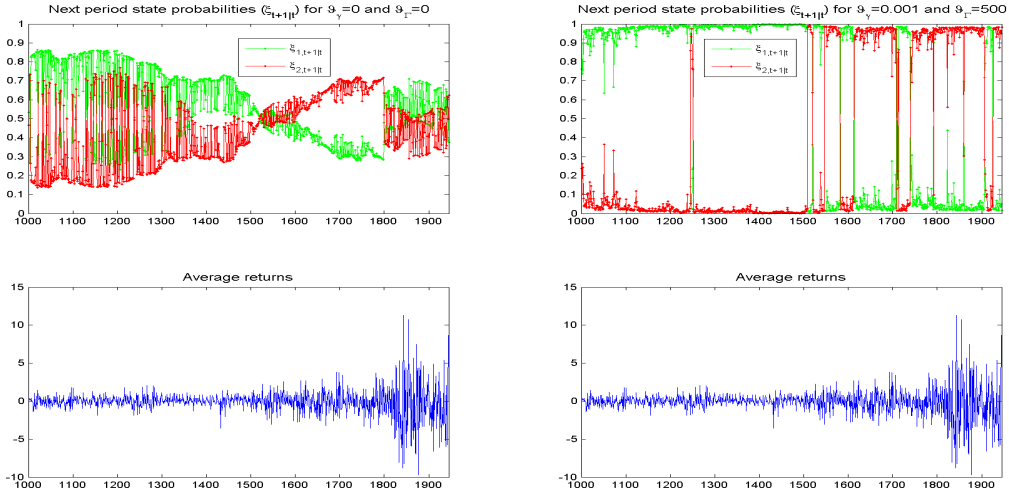

Figure 5.5: **Top:** Next period state probabilities over time for zero shrinkage case (left) and optimal shrinkage strength (right). **Bottom:** Average returns for $t = 1001, \ldots, 1945$.

based models. It is given by

$$\mathbf{\Sigma}_{t+d|t} = \mathbb{E}_{t+d|t}\left[G_{t+d}\right]\mathbb{E}_{t+d|t}\left[\mathbf{H}_{t+d|t}\right] + \mathbb{V}_{t+d|t}\left(G_{t+d}\right)\boldsymbol{\gamma}\boldsymbol{\gamma}',$$

where $\mathbb{E}_{t+d|t}\left[G_{t+d}\right] = \lambda$, $\mathbb{V}_{t+d|t}\left(G_{t+d}\right) = \lambda$ and $\mathbb{E}_{t+d|t}\left[\mathbf{H}_{t+d|t}\right] = \mathbf{S}_{t+d|t}\mathbb{E}_{t+d|t}\left[\mathbf{\Gamma}_{t+d|t}\right]\mathbf{S}_{t+d|t}$. Where $\mathbf{S}_{t+d|t}$ is formed by useing the respective volatility recursive equation, (5.6) or (5.5),

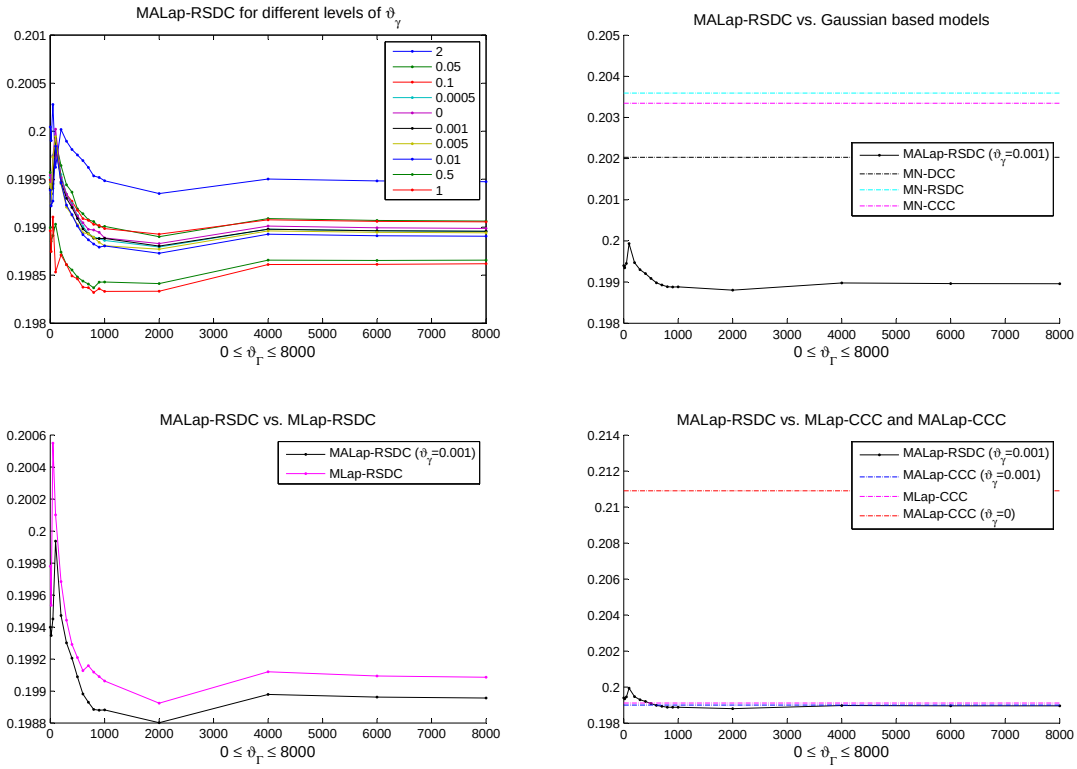

Figure 5.6: Comparison of the forecasting performance of the variance covariance matrix for different models. Setup similar to Figure 5.2, except that here, the RMSE measure is used.

and $\mathbb{E}_{t+d|t}\left[\mathbf{\Gamma}_{t+d|t}\right]$ reduces to $\mathbf{\Gamma}$ and $\sum_{n=1}^{N}\xi_{n,t+d|t}\mathbf{\Gamma}_n$ in the CCC and the RSDC models, respectively.

The cumulative predictive log-likelihood captures the accuracy of the whole shape of the distribution function. In order to empirically investigate the accuracy of the covariance matrix prediction, which results from our two step estimation approach, we compare forecasts of the covariance matrix from the aforementioned models. For this purpose, we employ two criteria which focus on the quality of the volatility forecasts (see Andersen et al., 1999; Pelletier, 2006)

$$\text{RMSE}_d = \left(\frac{1}{K^2}\sum_{i,j}\mathbb{E}\left[\left(\mathbf{\Sigma}_{i,j,t+d|t} - y_{i,t+d}y_{j,t+d}\right)^2\right]\right)^{1/2}, \tag{5.27}$$

$$\text{MAD}_d = \frac{1}{K^2}\sum_{i,j}\mathbb{E}\left[\left|\mathbf{\Sigma}_{i,j,t+d|t} - y_{i,t+d}y_{j,t+d}\right|\right], \tag{5.28}$$

where the MAD_d is based on absolute deviations and is considered being more robust to outliers.

Results are summarized in Figure 5.6 for the RMSE measure and in Figure 5.7 for the MAD measure. The MALap-RSDC model with shrinkage turns out to dominate all other models. The optimal level of shrinkage is higher in both cases than for the predictive log-likelihood measure.

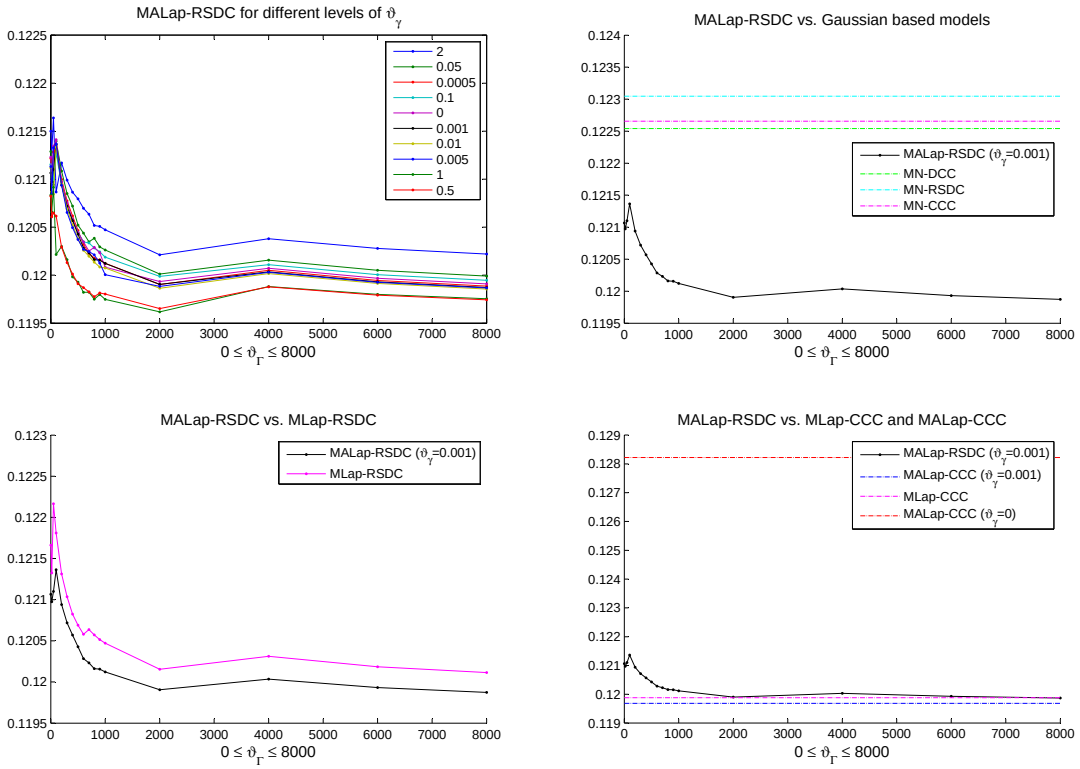

Figure 5.7: Comparison of the forecasting performance of the variance covariance matrix for different models. Setup similar to Figure 5.2, except that here, the MAD measure is used.

The MALap-RSDC is better than any Gaussian based model (top right panels in both figures) and with the shrinkage parameters $\boldsymbol{\vartheta_\gamma} = 0.001$ and $\boldsymbol{\vartheta_\Gamma} = 500$ it gives better predictions than the same model with the symmetric Laplace density and comparable results to a much simpler MLap-CCC model (bottom right panels in both figures).

The bottom right panel in Figure 5.7 indicates that the MALap-RSDC is slightly worst than the MLap-CCC (symmetric) model, but this measure is designed to be more robust to outliers which plays a crucial role in forming asymmetries in our data (or even in financial data in general). Hence, their impact should not be down weighted and we are not taking this result as a strong contradiction of our asymmetry assumption.

The empirical results show that any Laplace based model is strongly outperforming all the Gaussian based models, both in terms of the density prediction and the forecast of the matrix of second moments. Among all the Laplace based models the MALap-RSDC performs best and the forecast of the matrix of second moments is comparable to other Laplace based models for which the covariance matrices were estimated jointly with other parameters and by the one-step methods.

5.5 Conclusion

We extend the existing RSDC model of Pelletier (2006) to a more general class of distributions - the multivariate generalized hyperbolic class. By doing so we combine the model proposed in PP13 with a regime switching structure for the conditional dependency matrix. In order to maintain applicability of the proposed model for large number of assets, we propose a two-step estimation procedure. Further, we give the necessary and sufficient conditions for the consistency and asymptotic normality of the resulting two-step estimator.

In the empirical part, we investigate the performance of the new model and the two-step estimation method under Laplace distributed (special case of MGHyp) innovations. On the basis of in-sample fit and out-of-sample forecasting performance, all the Laplace based models clearly outperform standard Gaussian models from the literature like Constant Conditional Correlation (CCC) model of Bollerslev (1990), Dynamic Conditional Correlation (DCC) model of Engle and Sheppard (2002) and Regime Switching Conditional Correlation (RSDC) model of Pelletier (2006).

We show that the multivariate asymmetric Laplace model with a regime switching correlation matrix when combined with shrinkage has the best forecasting performance among all considered models. In particular, it beats all the Gaussian based models and also simpler Laplace based models. Our study confirms that both the heavy-tail assumption and three types of asymmetries in the distribution (due to the asymmetric distribution, due to the asymmetric response of the scale terms to the last period returns shock and due to the regime switching correlation structure) are important factors in the prediction of the distribution of assets returns.

It is important to emphasize that the model considered herein inherits the following important benefits of the original COMFORT model: (i) fast estimation because of the proposed two-step estimation procedure; (ii) the possibility of using parallel computing for further computational speed; (iii) the predictive distribution (here conditional on the next period regime) is a conditional MGHyp (or one of various special cases of it), so that, recalling that the sums of the margins of a MGHyp is itself a GHyp, portfolio construction and risk forecasting is straightforward. Future studies will include evaluation of the proposed models based on their portfolio performance and value-at-risk backtesting.

5.6 Appendices

5.6.1 Convergence of the Two-Step ECME Algorithm for the RSDC Model

We show below that, as the sample size goes to infinity, and under the (usual necessary) assumption that the model reflects the true DGP, except for the unknown parameters, then ignoring the correlation structure does not change estimates of $\boldsymbol{\theta}_P$ in the ECME algorithm. This result, together with some mild relative entropy condition given below, implies that the

monotonic increase of the incomplete-data likelihood is preserved when the ECME algorithm is performed on the subspace of the parameters space with zero correlations. Therefore, by the self consistency of the ECME algorithm, our approach maintains the consistency and asymptotic normality of estimates in the same sense as the existing two-step procedures under the normality assumption.

Since our ECME algorithm is iterating over zero correlation subspace of parameters space, the update of the information about $g_t^{\pm 1}$ in the E-step of Stage-I is incomplete, also the maximization in (5.15), on the zero correlation subspace, does not have to give the global optimum. Any version of the EM algorithm uses the monotonic convergence of the sequence of likelihood values $L_{\mathbf{Y}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C)$ based on a consecutive updates of all the parameters. We show that ignoring the correlation structure in the E-step of Stage-I and (5.15), under some mild information inequality condition, does not harm the monotonicity of our algorithm. For this purpose, denote by $\mathrm{KL}[P||Q]$ Kullback-Leibler divergence measure of the probability measure Q from the probability measure P. Where for two distributions P and Q of a continuous random variable, Kullback-Leibler divergence measure is defined by an integral

$$\mathrm{KL}[P||Q] = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x$$

where p and q denote the densities of P and Q, respectively.

Denote by $F_{G_t|\mathbf{Y}_t}(\cdot\,; \boldsymbol{\theta}_P, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C)$ the cumulative distribution function of $G_t \mid \mathbf{Y}_t$ with parameters $(\boldsymbol{\theta}_P, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C)$. The following theorem gives sufficient conditions for self consistency of the ECME algorithm performed on the zero correlation subspace.

Theorem 5.6.1 (Self consistency)**.** *Let $\boldsymbol{\theta}_C^*$ be the true value of parameters $\boldsymbol{\theta}_C$; ℓ and $\ell+1$ be the consecutive steps in the ECME algorithm performed on the zero correlation subspace $(\boldsymbol{\theta}_C = \mathbf{I}_K)$.*

If the inequalities

$$\begin{aligned}
&KL\Big[F_{G_t|\mathbf{Y}_t}\left(\cdot\,; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \mathbf{I}_K\right) || F_{G_t|\mathbf{Y}_t}\left(\cdot\,; \boldsymbol{\theta}_P^{[\ell+1]}, \boldsymbol{\theta}_D^{[\ell+1]}, \boldsymbol{\theta}_C^*\right)\Big] \\
&- KL\Big[F_{G_t|\mathbf{Y}_t}\left(\cdot\,; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \mathbf{I}_K\right) || F_{G_t|\mathbf{Y}_t}\left(\cdot\,; \boldsymbol{\theta}_P^{[\ell+1]}, \boldsymbol{\theta}_D^{[\ell+1]}, \mathbf{I}_K\right)\Big] \\
&\geq KL\Big[F_{G_t|\mathbf{Y}_t}\left(\cdot\,; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \mathbf{I}_K\right) || F_{G_t|\mathbf{Y}_t}\left(\cdot\,; \boldsymbol{\theta}_P^{[\ell+1]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*\right)\Big] \\
&- KL\Big[F_{G_t|\mathbf{Y}_t}\left(\cdot\,; \boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \mathbf{I}_K\right) || F_{G_t|\mathbf{Y}_t}\left(\cdot\,; \boldsymbol{\theta}_P^{[\ell+1]}, \boldsymbol{\theta}_D^{[\ell]}, \mathbf{I}_K\right)\Big]
\end{aligned} \tag{5.29}$$

hold, for $\ell = 1, 2, \ldots$, then the incomplete-data likelihood function at the true $\boldsymbol{\theta}_C^*$ does not decrease for each iteration of the ECME algorithm performed on the zero correlation subspace, i.e.

$$L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P^{[\ell+1]}, \boldsymbol{\theta}_D^{[\ell+1]}, \boldsymbol{\theta}_C^*\right) \geq L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}, \boldsymbol{\theta}_C^*\right).$$

Proof. Let the incomplete-data likelihood function be denoted by $L_{\mathbf{Y}}(\boldsymbol{\theta}_P, \boldsymbol{\theta}_D, \boldsymbol{\theta}_C)$. Then by a

Bayes rule and the complete log likelihood function decomposition (5.11)

$$\begin{aligned}\log L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C\right) &= \log L_{\mathbf{Y},G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C\right) - \log k_{G|\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C\right)\\ &= \log L^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right) + \log L^{\mathrm{Corr}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C\right)\\ &+ \log L_G\left(\boldsymbol{\theta}_D\right) - \log k_{G|\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C\right),\end{aligned} \tag{5.30}$$

where $k_{G|\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C\right)$ denotes the conditional probability density function of $G \mid \mathbf{Y}$.

We perform the ECME algorithm on the zero correlations subspace and in the E-step of the algorithm we take the conditional expectation under the probability measure given by last step estimates of $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_D$ and with ignored correlation structure. Therefore we start by taking the expectation of both sides of (5.30) with respect to the conditional distribution of G_t given $\boldsymbol{Y}_t$, using the ℓth fit of parameters $(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D)$, together with $\boldsymbol{\theta}_C = \mathbf{I}_K$. What we get is

$$\begin{aligned}\log L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C\right) &= \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[\log L^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right) + \log L^{\mathrm{Corr}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C\right) \mid \mathbf{Y}_t\right]\\ &+ \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[\log L_G\left(\boldsymbol{\theta}_D\right) - \log k_{G|\mathbf{Y}}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C\right) \mid \mathbf{Y}\right].\end{aligned} \tag{5.31}$$

We split the proof into two problems.

First, we show that $\boldsymbol{\theta}_P^{[\ell+1]}$ is chosen so that it maximizes not only

$$\mathcal{L}^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right) := \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[\log L^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right) \mid \mathbf{Y}_t\right]. \tag{5.32}$$

But also

$$\mathcal{L}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C\right) := \mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[\log L^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P\right) + \log L^{\mathrm{Corr}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C\right) \mid \mathbf{Y}_t\right]. \tag{5.33}$$

We do it by showing that both sets of first order conditions with respect to $\boldsymbol{\theta}_P$ are satisfied for the same vector of parameters. Hence the $\ell+1$ step estimates of $\boldsymbol{\theta}_P$ from the zero correlation model and the full model coincide.

$\boldsymbol{\theta}_P^{[\ell+1]}$ are chosen so that:

$$\frac{\partial \mathcal{L}^{\mathrm{MV}}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D\right)}{\partial \theta_{1,k,j}} = 0 \quad \Leftrightarrow$$

for $\theta_{1,k,j}$ equal ω_k, α_k or β_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[-\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \theta_{1,k,j}} + g_t^{-1}\left(Y_{k,t}-\mu_k-\gamma_k g_t\right)s_{k,t}^{-1}s_{k,t}^{-1}\left(Y_{k,t}-\mu_k-\gamma_k g_t\right)\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \theta_{1,k,j}}\right] = 0; \tag{5.34}$$

for $\theta_{1,k,j}$ equal μ_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[-\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \mu_k} + g_t^{-1}\left(Y_{k,t}-\mu_k-\gamma_k g_t\right)s_{k,t}^{-1}s_{k,t}^{-1}\left(\left(Y_{k,t}-\mu_k-\gamma_k g_t\right)\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \mu_k} - 1\right)\right] = 0; \tag{5.35}$$

for $\theta_{1,k,j}$ equal γ_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[-\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \gamma_k}+g_t^{-1}\left(Y_{k,t}-\mu_k-\gamma_k g_t\right)s_{k,t}^{-1}s_{k,t}^{-1}\left(\left(Y_{k,t}-\mu_k-\gamma_k g_t\right)\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \gamma_k}-g_t\right)\right]=0. \tag{5.36}$$

On the other hand, $\boldsymbol{\theta}_P^{[\ell+1]}$ which maximizes (5.33) must satisfy

$$\frac{\partial \mathcal{L}_{\mathbf{Y}|G}\left(\boldsymbol{\theta}_P,\boldsymbol{\theta}_C\right)}{\partial \theta_{1,k,j}}=0 \quad \Leftrightarrow$$

for $\theta_{1,k,j}$ equal ω_k, α_k or β_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[\begin{array}{c}-\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \theta_{1,k,j}}+g_t^{-1}\left(\mathbf{Y}_t-\boldsymbol{\mu}-\boldsymbol{\gamma}g_t\right)\mathbf{S}_t^{-1}\boldsymbol{\Gamma}_t^{-1}\left[\begin{array}{c}0\\ \vdots\\ \left(Y_{k,t}-\mu_k-\gamma_k g_t\right)s_{k,t}^{-1}\\ \vdots\\ 0\end{array}\right]\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \theta_{1,k,j}}\end{array}\right]=0; \tag{5.37}$$

for $\theta_{1,k,j}$ equal μ_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[\begin{array}{c}-\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \mu_k}+g_t^{-1}\left(\mathbf{Y}_t-\boldsymbol{\mu}-\boldsymbol{\gamma}g_t\right)\mathbf{S}_t^{-1}\boldsymbol{\Gamma}_t^{-1}\left(\left[\begin{array}{c}0\\ \vdots\\ \left(Y_{k,t}-\mu_k-\gamma_k g_t\right)s_{k,t}^{-1}\\ \vdots\\ 0\end{array}\right]\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \mu_k}-s_{k,t}^{-1}\right)\end{array}\right]=0; \tag{5.38}$$

for $\theta_{1,k,j}$ equal γ_k we have

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[\begin{array}{c}-\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \gamma_k}+g_t^{-1}\left(\mathbf{Y}_t-\boldsymbol{\mu}-\boldsymbol{\gamma}g_t\right)\mathbf{S}_t^{-1}\boldsymbol{\Gamma}_t^{-1}\left(\left[\begin{array}{c}0\\ \vdots\\ \left(Y_{k,t}-\mu_k-\gamma_k g_t\right)s_{k,t}^{-1}\\ \vdots\\ 0\end{array}\right]\frac{1}{s_{k,t}}\frac{\partial s_{k,t}}{\partial \gamma_k}-s_{k,t}^{-1}g_t\right)\end{array}\right]=0. \tag{5.39}$$

When $\theta_{1,k,j}$ equals ω_k, α_k or β_k, by applying the trace operator, we get that

$$g_t^{-1}\left(\mathbf{Y}_t-\boldsymbol{\mu}-\boldsymbol{\gamma}g_t\right)\mathbf{S}_t^{-1}\boldsymbol{\Gamma}_t^{-1}\left[0,\ldots,\left(Y_{k,t}-\mu_k-\gamma_k g_t\right)s_{k,t}^{-1},\ldots,0\right]'$$

is a random variable with the same mean as $g_t^{-1}\left(Y_{k,t}-\mu_k-\gamma_k g_t\right)s_{k,t}^{-1}s_{k,t}^{-1}\left(Y_{k,t}-\mu_k-\gamma_k g_t\right)$. Analogous result follows for $\theta_{1,k,j}$ equal to μ_k or γ_k.

From this we conclude that the value of $\boldsymbol{\theta}_P^{[\ell+1]}$ that maximizes (5.32) also maximizes (5.33). Second, in order to get the thesis we need to find $\boldsymbol{\theta}_D^{[\ell+1]}$ which satisfies

$$\begin{aligned}&\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[\log L_G\left(\boldsymbol{\theta}_D^{[\ell+1]}\right)-\log k_{G|\mathbf{Y}}\left(\boldsymbol{\theta}_P^{[\ell+1]},\boldsymbol{\theta}_D^{[\ell+1]},\boldsymbol{\theta}_C^*\right)\mid\mathbf{Y}_t\right]\\&\geq\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[\log L_G\left(\boldsymbol{\theta}_D^{[\ell]}\right)-\log k_{G|\mathbf{Y}}\left(G\mid\mathbf{Y};\boldsymbol{\theta}_P^{[\ell+1]},\boldsymbol{\theta}_D^{[\ell]},\boldsymbol{\theta}_C^*\right)\mid\mathbf{Y}\right],\end{aligned}\tag{5.40}$$

where $\boldsymbol{\theta}_C^*$ denotes the true value of $\boldsymbol{\theta}_C$. In the CM2-step we choose $\boldsymbol{\theta}_D^{[\ell+1]}$ which maximizes $\log L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P^{[\ell+1]},\boldsymbol{\theta}_D,\mathbf{I}_K\right)$, by assumption this maximizer satisfies inequality (5.29), hence it also must satisfy (5.40).

Therefore the monotonicity of the incomplete-data likelihood function after each iteration is preserved, i.e.

$$L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P^{[\ell+1]},\boldsymbol{\theta}_D^{[\ell+1]},\boldsymbol{\theta}_C^*\right)\geq L_{\mathbf{Y}}\left(\boldsymbol{\theta}_P^{[\ell]},\boldsymbol{\theta}_D^{[\ell]},\boldsymbol{\theta}_C^*\right)$$

holds.

□

In our case $F_{G_t|\mathbf{Y}_t}$ follows a GIG distribution, and whether the inequality, (5.29), always holds requires further investigation. The next remark tailors it to our distributional assumption.

Remark 5.6.1. *Let $\boldsymbol{\Gamma}_t$ be the true conditional correlation at time t; ℓ and $\ell+1$ be the consecutive steps in the ECME algorithm performed on the zero correlation subspace ($\boldsymbol{\theta}_C=\mathbf{I}_K$).*

For $(G_t\mid\mathbf{Y}_t;\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C)\sim GIG\big(\lambda-K/2,m_{\boldsymbol{\Gamma}_t}+\chi,\psi+\boldsymbol{\gamma}'\left(\mathbf{S}_t\boldsymbol{\Gamma}_t\mathbf{S}_t\right)^{-1}\boldsymbol{\gamma}\big)$, condition (5.29) is equivalent to

$$\Psi(\lambda^{[\ell+1]},\chi^{[\ell+1]},\psi^{[\ell+1]})\geq\Psi(\lambda^{[\ell]},\chi^{[\ell]},\psi^{[\ell]}),\quad\text{for }\ell=1,2,\ldots,\tag{5.41}$$

where $m_{\boldsymbol{\Gamma}_t}=\left(\mathbf{Y}_t-\boldsymbol{\mu}\right)'\left(\mathbf{S}_t\boldsymbol{\Gamma}_t\mathbf{S}_t\right)^{-1}\left(\mathbf{Y}_t-\boldsymbol{\mu}\right)$, function $\Psi(\lambda,\chi,\psi)$ is given by

$$\Psi(\lambda,\chi,\psi)=\frac{K_{\lambda-\frac{K}{2}}\left(\sqrt{\left(m_{\boldsymbol{\Gamma}_t}^{[\ell+1]}+\chi\right)\left(\boldsymbol{\gamma}^{[\ell+1]}(\mathbf{S}_t\boldsymbol{\Gamma}_t\mathbf{S}_t)^{-1}\boldsymbol{\gamma}^{[\ell+1]}+\psi\right)}\right)}{K_{\lambda-\frac{K}{2}}\left(\sqrt{\left(m_{\mathbf{I}_K}^{[\ell+1]}+\chi\right)\left(\boldsymbol{\gamma}^{[\ell+1]}(\mathbf{S}_t\mathbf{S}_t)^{-1}\boldsymbol{\gamma}^{[\ell+1]}+\psi\right)}\right)}\left(\sqrt{\frac{m_{\boldsymbol{\Gamma}_t}^{[\ell+1]}+\chi}{m_{\mathbf{I}_K}^{[\ell+1]}+\chi}\frac{\boldsymbol{\gamma}^{[\ell+1]}(\mathbf{S}_t\mathbf{S}_t)^{-1}\boldsymbol{\gamma}^{[\ell+1]}+\psi}{\boldsymbol{\gamma}^{[\ell+1]}(\mathbf{S}_t\boldsymbol{\Gamma}_t\mathbf{S}_t)^{-1}\boldsymbol{\gamma}^{[\ell+1]}+\psi}}\right)^{\lambda-\frac{K}{2}}$$

and $m_{\boldsymbol{\Gamma}_t}^{[\ell+1]}$ denotes $m_{\boldsymbol{\Gamma}_t}$ evaluated at $\boldsymbol{\theta}_P^{[\ell+1]}$ ($\ell+1$ step estimates of $\boldsymbol{\theta}_P$).

Hence, in order to guarantee the monotonic increase of the likelihood function, the CM-2 step maximization should be performed subject to the constrain (5.41). Unfortunately (5.41) cannot be integrated into the algorithm because it depends on the unknown true conditional correlation matrix $\boldsymbol{\Gamma}_t$. Therefore, at this point, we state (5.29), or equivalently (5.41), as an assumption.

Finally, note that under the Laplace distribution $(G_t\mid\mathbf{Y}_t;\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C)\sim\text{GIG}\big(\lambda-K/2,m_{\boldsymbol{\Gamma}_t},2+\boldsymbol{\gamma}'\left(\mathbf{S}_t\boldsymbol{\Gamma}_t\mathbf{S}_t\right)^{-1}\boldsymbol{\gamma}\big)$, and from the expression for the moments of the GIG random variable condition (5.41) in Remark 5.6.1 reduces to the following moment condition:

$$\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell+1]},\boldsymbol{\theta}_D^{[\ell]},\boldsymbol{\theta}_C^*\right)}\left[G_t^{\lambda^{[\ell+1]}-\lambda^{[\ell]}}\mid\mathbf{Y}_t\right]\geq\mathbb{E}_{\left(\boldsymbol{\theta}_P^{[\ell+1]},\boldsymbol{\theta}_D^{[\ell]},\mathbf{I}_K\right)}\left[G_t^{\lambda^{[\ell+1]}-\lambda^{[\ell]}}\mid\mathbf{Y}_t\right].\tag{5.42}$$

A consequence of Theorem 5.6.1 is that, if the starting values for the ECME algorithm, $\left(\boldsymbol{\theta}_P^{[1]}, \boldsymbol{\theta}_D^{[1]}\right)$, are sufficiently close to the true parameters, or if the log likelihood is unimodal in the parameters space, and condition (5.41) holds at each iteration of the algorithm, then monotonicity in the likelihood values of the estimates $\left(\boldsymbol{\theta}_P^{[\ell]}, \boldsymbol{\theta}_D^{[\ell]}\right)$ guarantees their convergence to the corresponding maximum likelihood estimates, even when we restrict the space over which we perform the iterations to the zero correlation subspace.

Thus, if our ECME algorithm converges to the global maximizers of the incomplete data likelihood function with respect to the $(\boldsymbol{\theta}_P, \boldsymbol{\theta}_D)$, and the correlations parameters $\boldsymbol{\theta}_C$ can be estimated consistently, then the asymptotic properties of the two-step estimation are described in the following theorem.

Theorem 5.6.2. *If the usual assumptions for the validity of maximum likelihood estimator are satisfied for both log likelihood functions, then the two-step estimates are consistent and asymptotically normal distributed with asymptotic covariance matrix* $\mathbf{V}$*, i.e.,*

$$\sqrt{T}\left(\begin{bmatrix}\hat{\boldsymbol{\theta}}_P\\ \hat{\boldsymbol{\theta}}_D\\ \hat{\boldsymbol{\theta}}_C\end{bmatrix} - \begin{bmatrix}\boldsymbol{\theta}_P\\ \boldsymbol{\theta}_D\\ \boldsymbol{\theta}_C\end{bmatrix}\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}),$$

with

$$\mathbf{V} = \begin{bmatrix}\mathbf{J}^{-1}_{\boldsymbol{\theta}_P,\boldsymbol{\theta}_D} & -\mathbf{J}^{-1}_{\boldsymbol{\theta}_C,\boldsymbol{\theta}_D}\mathbf{J}_{\boldsymbol{\theta}_P,\boldsymbol{\theta}_D}\mathbf{M}^{-1}\\ 0 & \mathbf{M}^{-1}\end{bmatrix} \mathbb{E}\left[\frac{\partial \log f}{\partial(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C)}\frac{\partial \log f}{\partial(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C)}\right] \begin{bmatrix}\mathbf{J}^{-1}_{\boldsymbol{\theta}_P,\boldsymbol{\theta}_D} & -\mathbf{J}^{-1}_{\boldsymbol{\theta}_P,\boldsymbol{\theta}_D}\mathbf{J}_{\boldsymbol{\theta}_P,\boldsymbol{\theta}_D}\mathbf{M}^{-1}\\ 0 & \mathbf{M}^{-1}\end{bmatrix},$$

where

$$\mathbf{J}_{\boldsymbol{\theta}_P,\boldsymbol{\theta}_D} = \mathbb{E}\left[\frac{\partial g(\mathbf{Y}_t,\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C)}{\partial(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D)}\right],\ \mathbf{J}_{\boldsymbol{\theta}_C} = \mathbb{E}\left[\frac{\partial g(\mathbf{Y}_t,\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C)}{\partial\boldsymbol{\theta}_C}\right],\ \mathbf{M} = \mathbb{E}\left[\frac{\partial m(\mathbf{Y}_t,\boldsymbol{\theta}_C)}{\partial\boldsymbol{\theta}_C}\right],$$

$$g(\mathbf{Y}_t,\boldsymbol{\theta}_P,\boldsymbol{\theta}_D,\boldsymbol{\theta}_C) = \frac{\partial \log f(\mathbf{Y}_t \mid \mathbf{Y}_{t-1},\ldots,\mathbf{Y}_1)}{\partial(\boldsymbol{\theta}_P,\boldsymbol{\theta}_D)}, \qquad m(\mathbf{Y}_t,\boldsymbol{\theta}_C) = \frac{\partial \log f(\mathbf{Y}_t \mid \mathbf{Y}_{t-1},\ldots,\mathbf{Y}_1)}{\partial\boldsymbol{\theta}_C}.$$

The matrix $\mathbf{V}$ *can be consistently estimated by its plug-in estimates.*

Proof: see Newey and McFadden (1994).

5.6.2 DJ-30 Components

Dow Jones Industrial Average components (until June 8, 2009)

No.	Abbrev	Company Name	No.	Abbrev	Company Names
1	MMM	3M Co	16	HD	Home Depot Inc
2	AA	Alcoa Inc	17	IBM	Intl. Business Machines Corp
3	AXP	American Express Co	18	INTC	Intel Corp
4	T	AT& T	19	JNJ	Johnson & Johnson
5	BAC	Bank of America Corp	20	JPM	JPMorgan Chase & Co
6	BA	Boeing Co	21	KFT	Kraft Foods Inc
7	CAT	Caterpillar Inc	22	MCD	McDonald's Corp
8	CVX	Chevron Corp	23	MRK	Merck & Co Inc
9	C	Citigroup Inc	24	MSFT	Microsoft Corp
10	KO	Coca-Cola Co	25	PFE	Pfizer Inc
11	DD	El Du Pont de Nemours & Co	26	PG	Procter & Gamble Co
12	XOM	Exxon Mobil Corp	27	UTX	United Technologies Corp
13	GE	General Electric Co	28	VZ	Verizon communications Inc
14	GM	General Motors Corp	29	WMT	Wal-Mart Stores Inc
15	HPQ	Hewlett-Packard Co	30	DIS	Walt Disney Co

Bibliography

Aas, K. and Haff, I. H. (2006a). The Generalized Hyperbolic Skew Students t-Distribution. *Journal of Financial Econometrics*, 4(2):275–309. 73

Aas, K. and Haff, I. H. (2006b). The Generalized Hyprbolic Skew Student's t-Distribution. *Journal of Fiancial Econometrics*, 4(2):275–309. 23

Aielli, G. P. (2011). Dynamic Conditional Correlation: On Properties and Estimation. Available at ssrn: http://ssrn.com/abstract=1507743 or http://dx.doi.org/10.2139/ssrn.1507743. 2, 31, 47, 50, 51

Aït-Sahalia, Y., Cacho-Diaz, J., and Hurd, T. R. (2009). Portfolio Choice With Jumps: A Closed-Form Solution. *The Annals of Applied Probability*, 19(2):556–584. 11

Aït-Sahalia, Y. and Jacod, J. (2011). Testing Whether Jump Have Finite or Infinite Activity. *The Annals of Statistics*, 39(3):1689–1719. 12

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *The Review of Economics and Statistics*, 89(4):701–702. 11

Andersen, T. G., Bollerslev, T., and Lange, S. (1999). Forecasting Financial Market Volatility: Sample Frequency vis-a-vis Forecast Horizon. *Journal of Empirical Finance*, 6(5):457–477. 87

Asai, M. and McAleer, M. (2009). The Structure of Dynamic Correlations in Multivariate Stochastic Volatility Models. *Journal of Econometrics*, 150:182–192. 11

Asai, M. and McAleer, M. (2011). Alternative Asymmetric Stochastic Volatility Models. *Econometric Reviews*, 30(5):548–564. 16

Asai, M., McAleer, M., and Yu, J. (2006). Multivariate Stochastic Volatility: A Review. *Econometric Reviews*, 25:145–175. 11

Barndorff-Nielsen, O. E. (1978). Hyperbolic Distributions and Distributions on Hyperbolae. *Scandinavian Journal of Statistics*, 5:151–157. 6

Barndorff-Nielsen, O. E. (1997). Processes of Normal Inverse Gaussian Type. *Finance and Stochastics*, 2:41–68. 22

Barone-Adesi, G., ngle, R. F., and Mancini, L. (2008). A GARCH Option Pricing Model with Filtered Historical Simulations. *The Review of Financial Studies*, 21(3):1223–1258. 23

Bickel, P. J. and Levina, E. (2008). Regularized Estimation of Large Covariance Matrices. *The Annals of Statistics*, 36(1):199–227. 5

Blæsid, P. and Sørensen, M. (1992). 'HYP'-a Computer Program for Analyzing Data by Means of the Hyperbolic Distribution. Technical report, Research Report 248, Department of Theoretical Statistics, Aarhus University. 7

Bollerslev, T. (1990). Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach. *The Review of Economics and Statistics*, 72:498–505. 2, 3, 18, 19, 31, 46, 72, 76, 81, 89

Bollerslev, T., Law, T. H., and Tauchen, G. (2008). Risk, Jumps, and Diversification. *Journal of Econometrics*, 144(1):234–256. 11

BIBLIOGRAPHY

Bollerslev, T., Todorov, V., and Li, S. Z. (2013). Jump tails, extreme dependencies, and the distribution of stock returns. *Journal of Econometrics*, 172(2):307–324. 11

Bos, C. S. (2012). chapter Relating Stochastic Volatility Estimation Methods, pages 147–175. John Wiley & Sons Inc. 11

Broda, S. A., Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2012). Stable Mixture GARCH Models. *Journal of Econometrics*, 172(2):292306. 5, 74

Broda, S. A. and Paolella, M. S. (2009). CHICAGO: A Fast and Accurate Method for Portfolio Risk Calculation. *Journal of Financial Econometrics*, 1:1–25. 6, 22

Broda, S. A. and Paolella, M. S. (2010). Saddlepoint Approximation of Expected Shortfall for Transformed Means. *UvA Econometrics Discussion Paper 2010/08. University of Amsterdam.* 54

Broda, S. A. and Paolella, M. S. (2011). Expected Shortfall for Distributions in Finance. In Čížek, P., Härdle, W., and Rafał Weron, editors, *Statistical Tools for Finance and Insurance*. 54

Caporin, M. and McAleer, M. (2013). Ten Things You Should Know about the Dynamic Conditional Correlation Representation. *Econometrics*, 1(1):115–126. 47

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2002). The Fine Structure of Asset Returns: An Empirical Investigation. *The Journal of Business*, 75(2):pp. 305–333. 12

Chan, W. H. and Maheu, J. M. (2002). Conditional Jump Dynamics in Stock Market Returns. *Journal of Businesss and Economic Statistics*, 20(3):377–389. 11, 14

Chernov, M., Gallant, A. R., Ghysels, E., and Tauchen, G. (2003). Alternative Models for Stock Price Dynamics. *Journal of Econometrics*, 116(3):225–257. 11

Chopra, V. and Ziemba, W. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19:6–12. 35

Christoffersen, P., Elkamhi, R., Feunou, B., and Jacobs, K. (2010). Option Valuation with Conditional Heteroskedasticity and Nonnormality. *The Review of Financial Studies*, 23(5):2139–2183. 23, 24, 39

Christoffersen, P., Heston, S., and Jacobs, K. (2006). Option Valuation with Conditional Skewness. *Journal of Econometrics*, 131:253–284. 23

Diebold, F. X. and Mariano, R. S. (2002). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144. 34

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, 1:83–106. 74

Doganoglu, T., Hartz, C., and Mittnik, S. (2007). Portfolio Optimization when Risk Factors are Conditionally Varying and Heavy Tailed. *Computational Economics*, 29(3-4):333–354. 6

Duan, J.-C. (1995). The GARCH Option Pricing Model. *Mathematical Finance*, 5(1):13–32. 23

Eberlein, E. and Keller, U. (1995). Hyperbolic Distributions in Finance. *Bernoulli*, 1:281–299. 6, 12, 13

Eberlein, E., Keller, U., and Prause, K. (1998). New Insights into Smile, Mispricing, and Value at Risk: the Hyperbolic Model. *Journal of Business*, 38:371–405. 6, 13

Embrechts, P., McNeil, A. J., and Straumann, D. (2001). *Risk Management: Value at Risk and Beyond*, chapter Correlation and Dependency in Risk Management: Properties and Pitfalls, pages 176–223. Cambridge University Press, Cambridge. 12

Engle, R. (2009). *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton University Press, Princeton. 46

Engle, R. F. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics*, 20:339–350. 2, 19, 31, 38, 46, 50, 51

Engle, R. F. and Ng, V. K. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, 48(5):pp. 1749–1778. 15

Engle, R. F. and Sheppard, K. (2002). Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. SSRN-id:1296441. 18, 81, 89

Eraker, B. (2004). Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. *Journal of Finance*, 59(3):1367–1403. 11

Eraker, B., Johannes, M., and Polson, N. (2003). The Impact of Jumps in Volatility and Returns. *Journal of Finance*, 58(3):1269–1299. 11

Fan, J., Fan, Y., and Lv, J. (2008). High Dimensional Covariance Matrix Estimation Using a Factor Model. *Journal of Econometrics*, 147:186–197. 5

Francq, C. and Zakoïan, J.-M. (2004). Maximum Likelihood estimation of Pure GARCH and ARMA-GARCH Processes. *Bernoulli*, 10(4):605–637. 5

Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models Structure, Statistical Inference and Financial Applications*. John Wiley & Sons Ltd. 5

Giacometti, R., Bertocchi, M. I., Rachev, S. T., and Fabozzi, F. J. (2007). Stable Distributions in the Black-Litterman Approach to Asset Allocation. *Quantitative Finance*, 7(4):423–433. 6

Giacomini, R. and White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica*, 74(6):1545–1578. 34

Gilder, D., Shackleton, M. B., and Taylor, S. J. (2013). Cojumps in Stock Prices: Empirical Evidence. Accepted Manuscript, doi: http://dx.doi.org/10.1016/j.jbankfin.2013.04.025. 11

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, 48(5):pp. 1779–1801. 15, 49

Greenspan, A. (1997). Maintaining Financial Stability in a Global Economy. Discussion at Federal Reserve Bank of Kansas City symposium. 1

Haas, M., Mittnik, S., and Paolella, M. S. (2004). Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics*, 2(2):211–250. 47

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57:257–84. 3, 79

Hamilton, J. D. (1991). A Quasi-Bayesian Approach to Estimating Parameters for Mixtures of Normal Distributions. *Journal of Business and Economic Statistics*, 9(1):27–39. 77

Hamilton, J. D. (1993). *Handbook of Statistics Vol. 11*, chapter Estimation, Inference, and Forecasting of Time Seris Subject to Changes in Regime. New York: North-Holland. 77

Hamilton, J. D. (1994). *Time Series Analysis*. Princton University Press. 77

Hansen, J. V., McDonald, J. B., and Turley, R. S. (2006). Partially Adaptive Robust Estimation of Regression Models and Applications. *European Journal of Operational Research*, 170:132–143. 6

Heston, S. L. and Nandi, S. (2000). A Closed-Form GARCH Option Valuation Model. *The Review of Financial Studies*, 13(3):585–625. 23

Jensen, M. B. and Lunde, A. (2001). The NIG-S&ARCH Model: A Fat-Tailed, Stochastic and Autoregressive Conditional Heteroskedastic Volatility Model. *Econometrics Journal*, 4:319–342. 73

Jondeau, E. (2012). Asymmetry in Tail Dependence of Equity Portfolios. Working Paper, HEC, University of Lausanne. 23, 38, 47

Jondeau, E. and Rockinger, M. (2005). Conditional asset allcoation under non-normality how costrly is the mean-variance criterion? *Moscow Meetings Discussion Paper*, EFA. 47

Kim, C.-J. (1994). Dynamic Linear Models with Markov-Switching. *Journal of Econometrics*, 60(1-2):1–22. 77, 79

BIBLIOGRAPHY

Kotz, S., Podgórski, K., and Kozubowski, T. (2001). *The Laplace Distribution and Generalizations: A Revist with Application to Communication, Economics, Engineering and Finance*. Massachusetts: Birkhäuser. 22

Kotz, S., Podgórski, T. J., and Kozubowski, T. (2000). An Asymmetric Multivariate Laplace Distribution. Technical report, Department of Statistics and Applied Probability, University of California at Santa Barbara. 22

Kozubowski, T. J. and Podgórski, K. (2001). Asymmetric Laplace Laws and Modeling Financial Data. *Mathematical and Computer Modelling*, 34:1003–1021. 22

Kuester, K., Mittik, S., and Paolella, M. S. (2006a). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4(1):53–89. 63

Kuester, K., Mittnik, S., and Paolella, M. S. (2006b). Value–at–Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, 4:53–89. 5

Liu, C. and Rubin, D. B. (1994). The ECME algorithm: A Simple Extension of EM and ECM with Faster Monotone Convergence. *Biometrika*, 81, No. 4:633–648. 18

Madan, D. B., Carr, P. P., and Chang, E. C. (1998). The Variance Gamma Process and Option Pricing. *European Finance Review*, 2(1):79–105. 12

Madan, D. B. and Seneta, E. (1990). The Variance Gamma (V.G.) Model for Share Market Returns. *The Journal of Business*, 63(4):pp. 511–524. 12, 22

Maheu, J. M. and McCurdy, T. H. (2004). News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns. *Journal of Finance*, 59(2):755–793. 11, 14

Markowitz, H. (1952). Portfolio Selection. *Jornal of Finance*, 7:77–91. 57

McDonald, J. B. (1991). Parametric Models for Partially Adaptive Estimation with Skewed and Leptokurtic Residuals. *Economics Letters*, 37:273–278. 6

McDonald, J. B. (1997). Probability Distributions for Financial Models. In Maddala, G. S. and Rao, C. R., editors, *Handbook of Statistics*, volume 14. Elsevier Science. 6

McDonald, J. B. and Newey, W. K. (1988). Partially Adaptive Estimation of Regression Models via the Generalized t Distribution. *Econometric Theory*, 4:428–457. 6

McKay, A. T. (1932). A bessel function distribution. *Biometrika*, 24(1/2):pp. 39–44. 22

McLachlan, G. J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*. Hoboken, New Jersey, second edition. 20

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princton University Press. 6, 7, 13, 16, 22, 23, 58, 73

Mencía, J. and Sentana, E. (2009). Multivariate Location-Scale Mixtures of Normals and Mean-Variance-Skewness Portfolio Allocation. *Journal of Econometrics*, 153(2):105–121. 7, 16, 52, 54

Mittnik, S. and Paolella, M. S. (2000). Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates. *Journal of Forecasting*, 19:313–333. 15, 63, 74

Newey, W. K. and McFadden, D. (1994). *Large sample estimation and hypothesis testing. In: Handbook of Econometrics vol. IV*, chapter 36, pages 2111–2245. North-Holland. 94

Noureldin, D., Shephard, N., and Sheppard, K. (2012). Multivariate High-Frequency-Based Volatility (HEAVY) Models. *Journal of Applied Econometrics*, 27(6):907–933. 38

Paolella, M. S. (2007). *Intermediate Probability: A Computational Approach*. Wiley-Interscience. 6, 21, 22, 42

Paolella, M. S. (2013). Multivariate Asset Return Prediction with Mixture Models. Forthcoming in: European Journal of Finance. 6, 19, 21, 32, 35, 47, 53, 72, 77, 82, 83

Paolella, M. S. and Polak, P. (2013a). ALRIGHT: Asymmetric LaRge-Scale (I)GARCH with Hetero-Tails. 20, 47, 51

Paolella, M. S. and Polak, P. (2013b). COMFORT: A Common Market Factor Non-Gaussian Returns Model. Research Paper 13-38, Swiss Finance Institute, Available at SSRN: http://ssrn.com/abstract=2287793 or http://dx.doi.org/10.2139/ssrn.2287793. 34, 47, 73

Pelletier, D. (2006). Regime Switching for Dynamic Correlations. *Journal of Econometrics*, 131:445–473. 3, 18, 19, 38, 72, 76, 79, 81, 87, 89

Phillips, R. F. (1994). Partially Adaptive Estimation via a Normal Mixture. *Journal of Econometrics*, 64:123–144. 6

Prause, K. (1999). *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*. PhD thesis, University of Freiburg. 7, 22

Protassov, R. (2004). EM-based Maximum Likelihood Parameter Estimation for Multivariate Generalized Hyperbolic Distributions with Fixed Lambda. *Statistics and Computing*, 14:1:67–77. 7, 16, 21

Rockafeller, R. T. and Uryasev, S. (2000). Optimization of Conditional Value at Risk. *Journal of Risk*, 2:21–41. 57

Rombouts, J. V. and Stentoft, L. (2011). Multivariate Option Pricing with Time Varying Volatility and Correlations. *Journal of Banking and Finance*, 35:2267–2281. 23, 24, 39

Scott, D. J., Würtz, D., Dong, C., and Tran, T. T. (2011). Moments of the Generalized Hyperbolic Distribution. *Computational Statistics*, 26:459–476. viii, 16, 30, 32, 52, 55

Taylor, S. J. (1982). Financial Returns Modelled by the Product of Two Stochastic Processes - a Study of Daily Sugar Prices 1961-79. In *Time Series Analysis: Theory and Practice, 1*, pages 203–226. Amsterdam: North-Holland. 11, 14, 49

Todorov, V. and Tauchen, G. (2011). Volatility Jumps. *American Statistical Association*, 29(3):356–371. 11

Tse, Y. K. and Tsui, A. K. C. (2002). A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations. *Journal of Business and Economic Statistics*, 20(3):351–362. 2, 31, 38, 46, 50

Watson, G. N. (1922). *A Treatise on the Theory of Bessel Functions*. 43